

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2001, 10:47:22 ; Search time 43.71 Seconds

(without alignments)
5820.960 Million cell updates/sec

Title: US-09-596-958-1

Perfect score: 1344

Sequence: 1 atgtcaattcttcgcttaa.....acctgaaggctgctgaatga 1344

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	155.2	11.5	1729	4	US-09-120-817-1
2	94.8	7.1	591	4	US-09-402-668-1
3	74.2	5.5	666	4	US-09-198-956-3
4	58.6	4.4	185	4	US-09-402-668-9
5	45.8	3.4	7218	1	US-08-232-463-14
6	43.2	3.2	7218	1	US-08-232-463-14
7	37.4	2.8	2371	2	US-08-343-443B-1
8	36.2	2.7	400	3	US-08-961-083-187
9	35.8	2.7	53526	3	US-08-658-136-2
10	35.8	2.7	53577	3	US-08-658-136-1
11	35.2	2.6	1341	2	US-08-945-848-7
12	35.2	2.6	2363	2	US-08-945-848-6
13	34.6	2.6	1931	2	US-09-130-114-2
14	34.2	2.5	485	5	PCT-US95-05741-10
15	34	2.5	456	1	US-08-233-789A-58
16	33.2	2.5	1849	2	US-08-676-166A-1
17	32.8	2.4	1833	2	US-08-403-852D-6
18	32.8	2.4	1833	3	US-08-510-646B-6
19	32.8	2.4	1833	3	US-09-231-818-6
20	32.4	2.4	3211	2	US-08-574-959A-8
21	32.4	2.4	3901	2	US-08-574-959A-6
22	32.2	2.4	1709	2	US-09-010-398-2
23	32.2	2.4	1709	4	US-09-366-260-2
24	32.2	2.4	1709	4	US-09-153-804-11
25	32	2.4	513	1	US-08-217-327-3
26	32	2.4	913	1	US-07-885-970A-3
27	32	2.4	913	1	US-08-298-687A-3

28	32	2.4	913	1	US-08-530-797-2	Sequence 2, Appl1
29	32	2.4	913	1	US-08-298-829-3	Sequence 3, Appl1
30	32	2.4	913	2	US-08-787-335-2	Sequence 2, Appl1
31	32	2.4	1984	1	US-07-885-970A-25	Sequence 25, Appl1
32	32	2.4	1985	1	US-08-298-687A-25	Sequence 25, Appl1
33	32	2.4	1985	1	US-08-298-829-25	Sequence 25, Appl1
34	32	2.4	3708	1	US-08-185-232A-1	Sequence 1, Appl1
35	32	2.4	3708	1	US-08-416-523-1	Sequence 1, Appl1
36	32	2.4	3708	3	US-08-789-478-1	Sequence 1, Appl1
37	32	2.4	4481	4	US-09-041-886-18	Sequence 18, Appl1
38	31.8	2.4	936	4	US-09-179-558-62	Sequence 62, Appl1
39	31.8	2.4	24417	2	US-08-846-762-1	Sequence 1, Appl1
40	31.6	2.4	1820	1	US-08-173-508-7	Sequence 7, Appl1
41	31.6	2.4	1821	2	US-08-265-310-7	Sequence 7, Appl1
42	31.6	2.4	1821	3	US-08-951-742-7	Sequence 7, Appl1
43	31.2	2.3	2335	4	US-09-387-574-9	Sequence 9, Appl1
44	31.2	2.3	2823	1	US-08-398-008A-1	Sequence 1, Appl1
45	31.2	2.3	2823	2	US-08-893-333-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-120-817-1
; Sequence 1, Application US/09120817
; Patent No. 6172184
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Charkowski, Amy
; APPLICANT: Alfano, James R.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: PSEUDOMONAS SYRINGAE AND ITS USE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/120,817
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,107
; FILING DATE: 06-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1741
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1729 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-120-817-1

Query Match 11.5%; Score 155.2; DB 4; Length 1729;
Best Local Similarity 57.1%; Pred. No. 1e-38;
Matches 330; Conservative 0; Mismatches 233; Indels 15; Gaps 2;

QY 745 ggcgtatcatcagacggtgtcgtcatcacaccattaccgtgaaagcgggtcaggtgtttgat 804
Db 1079 GCCGCAAGATCAATGTGTGTGAAGACACCACTCAAGTCCGGCTGGCGAACTTTTGAC 1138
QY 805 ggcgaagagacaacccctccagccggttcagaaattagcgtatggcggcagctctgaaac 864
Db 1139 GCCCAGCGCGCAACCTTCACTCTGCAACAAATATATGGTTAAGCGAGACCGGCGAAAT 1198
QY 865 cagaacccgcgtgttatactactgaagacggtgccagccctgaaacagctcaccatggcgac 924
Db 1199 CAGAAGCCCATGTTTCAGCTGCTGAGCGCTACGTGTGAAGAAATGTGAACCTGGGTGAG 1258
QY 925 gacggggcggtggtattcatctttacg-----gtgatgccaaaaatagacaa 972
Db 1259 AACGAGGTCGATGGCATCCACGTGAAGCAAAACGCTCAGGAAGTCACCATTCACAAC 1318
QY 973 ctgcacgtcacaacagctgggtgagacggtattaccgttaagccaaacagcggcgga 1032
Db 1319 GTGCATGCCCAAGACGTCGGTGAAGACCTGATTACGGTCAAGAGCGAGGCGCAGCG 1378
QY 1033 aaatcccgctgaaatcaactaacagttccttcgagcgcctctgacaagatcctcgag 1092
Db 1379 GTCACTAATCTCAACATCAAGAACAGCAGTGCCTAAGGTGCAGACGACAAAGTTGTCAG 1438
QY 1093 ctgaatgcgtactaaacctgagcgttgacaacgtgaaagcgaagactttggtactttt 1152
Db 1439 CTCACGCGCAACACTCACTTGAAATCGACAACACTTCAAGCGCGACGATTTCCGGCAGCATG 1498
QY 1153 gtacgcactaacggcggtcaacag--ggttaactgggtatcgaatcgtagccataatcgc 1209
Db 1499 GTTCGACCAACGGTGGCAAGCATTTGATGACATGAGCATCGAGCTGAACGGCATCGAA 1558
QY 1210 gcagagacggttaagtctctgtctgttaaaagcgtatagcggggttaagctcaatc 1269
Db 1559 GCTAACCGGCAAGTTGCGCTGTGTGAAGCGACAGTGCAGTCTGAGCTGACGATCTGAAGCTGCAACG 1618
QY 1270 agtgatatctcaactgggtggtgattgtgaaacacactacaa 1307
Db 1619 GGAACATCGCCATGACCGACGCTCAACACAGCGCTACGA 1656

RESULT 2

US-09-402-668-1
; Sequence 1, Application US/09402668
; Patent No. 6172030

GENERAL INFORMATION:

; APPLICANT: WADA, Yasunao
; APPLICANT: KASAI, Miyuki
; APPLICANT: SHIKATA, Shitsuw
; APPLICANT: SUZUMATSU, Atsushi
; APPLICANT: KOIKE, Kenzo
; APPLICANT: HATADA, Yuji
; APPLICANT: KOBAYASHI, Tohru
; APPLICANT: ITO, Susumu
; APPLICANT: TSUMADORI, Masaki
; TITLE OF INVENTION: Detergent Composition
; FILE REFERENCE: 2173-0116P
; CURRENT APPLICATION NUMBER: US/09/402,668
; CURRENT FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 9-091142 JAPAN
; PRIOR FILING DATE: 1997-04-09
; PRIOR APPLICATION NUMBER: 9-242736 JAPAN
; PRIOR FILING DATE: 1997-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/01613
; PRIOR FILING DATE: 1998-04-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Bacillus sp.

FEATURE:

OTHER INFORMATION: Strain: KSM-P15

; NAME/KEY: CDS
; LOCATION: (1)..(591)
US-09-402-668-1

Query Match 7.1%; Score 94.8; DB 4; Length 591;
Best Local Similarity 52.8%; Pred. NO. 4.8e-20;
Matches 263; Conservative 0; Mismatches 217; Indels 18; Gaps 2;

QY 754 cagacggtgctgcatgacacaccattaccgtgaaagcgggtcaggtgtttgatggcaagaa 813
Db 4 cagacggtgctgcatgaaacattcgtgctgcgtgcagacgtttgacggaaaggg 63
QY 814 caaaccttcaacgcgcggttcaga---attagcgatggcgccagctctgaaacacagaaa 870
Db 64 cagacctaagtggctaatccgaatacatatgggggacggtcgcagcgagagaaatcagaag 123
QY 871 cgcgtgtttatactggaagacggtgcccagcctgaaaaacgtccaccatggggacgcggg 930
Db 124 cgcattctcgtctgagggctggggcaagcctgaaaaatgtagtgattggcgctcctgcc 183
QY 931 gcggatggtattcatctttagcgtgtagtgcacaaatagacaaatctgcagtcaccaacgtg 990
Db 184 gctgacggggtgcactgctgacgggtgtagcattcaaatgctcatctggtggagatgtt 243
QY 991 ggtgagggcgcggtattacggttaagccaaacagcgcgggcaaaaaatcccacgttgaaatc 1050
Db 244 ggtgaggtgctgacgcttaaatcgtccggaacg-----gtgaacatc 288
QY 1051 actaacagtctcctgagcagcgcctctgacaagatcctgcagctgaatgcogataactaac 1110
Db 289 tcggcgggggcagcctacaaagcgtatgcaagggttcccaaatcaatgcagcggggagcg 348
QY 1111 ctgagcgttgcaaacgtgaaagcgaagacactttggtactttgtacgcactaacgcggt 1170
Db 349 atcaacattctaaactcaggcgcggtgacatcgggaagcgtgttcgcgcacaaagcggcg 408
QY 1171 caacagggtaactgggtatcgaatctgagccatatcagcgcagaaagacggtlaagtctcgc 1230
Db 409 accacctacaaagtggtgatgaacgtggaacactggaacatttccagagtgaagatcgcg 468
QY 1231 ttcgttaaaagcgatagc 1248
Db 469 atcctgagaacggacagc 486

RESULT 3

US-09-198-956-3
; Sequence 3, Application US/09198956
; Patent No. 6165769

GENERAL INFORMATION:

; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 666
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-198-956-3

Query Match	5.58;	Score 74.2;	DB 4;	Length 666;
Best Local Similarity	55.78;	Prod. NO. 1.5e-13;		
Matches 142;	Conservative	0;	Mismatches 113;	Indels 0; Gaps 0;
759 ggtgctgcgatgacaccattaccgtgaaacggtccaggtctgttgatggcaaaaggacaaac	818			
90 ggtcgttcacaaaacgatcgtagtcgagaaggccaaacgtacgacgaaaggcaagcg	149			
819 ctccacgcgcggttcagaattaggcgtatggcgcgcagctctgaaacacgaacacgcgtgtt	878			
150 gctgattcaggttcggcgtatcggggcgcgcgaacacgcgagga tcaaaaacccgatttt	209			
879 tatactggaagacggtgcacgcttgaaaaacgtccaccatggcgacgcggggcgatgg	938			
210 caaagtggaggatggtgcaacgcgtcaaaaatctgctgcttggcgctcctgctgctgatgg	269			
939 tattcatctttacggtgatgcctcaaaaatagacaattctgcacgtcaccacacgttggtgaga	998			
270 tgttcacacatattggaaacgctttccataaacacgttgtttggggaagatgtcgcggaaga	329			
999 cgcgattaccggttaa	1013			
330 tgccttgacttcaa	344			

```

RESULT      4
US-09-402-668-9
US-09-402-668-9, Application US/09402668
; Patent No. 6172030
; GENERAL INFORMATION:
; APPLICANT: WADA, Yasunao
; APPLICANT: KASAI, Miyuki
; APPLICANT: SHIKATA, Shitsuw
; APPLICANT: SUZUMATSU, Atsushi
; APPLICANT: KOIKE, Kenzo
; APPLICANT: HATADA, Yuji
; APPLICANT: KOBAYASHI, Tohru
; APPLICANT: ITO, Susumu
; APPLICANT: TSUMADORI, Masaki
; TITLE OF INVENTION: Detergent Composition
; FILE REFERENCE: 2173-0116P
; CURRENT APPLICATION NUMBER: US/09/402,668
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 9-091142 JAPAN
; PRIOR FILING DATE: 1997-04-09
; PRIOR APPLICATION NUMBER: 9-242736 JAPAN
; PRIOR FILING DATE: 1997-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/01613
; PRIOR FILING DATE: 1998-04-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 185
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:primer
US-09-402-668-9

```

	Query Match	4.4%	Score 58.6	DB 4	Length 185
	Best Local Similarity	63.3%	Pred. No. 5.3e-09		
	Matches 107	Conservative	0	Mismatches 59	Indels 3
	Gaps				
Qy	757	acggtgctgcatacacaccattacgctgaaacgcgggtcagtggtttgatggcgaaggacaa	816		
Db	7	actgctgcacgaacgattctgtcctgcgcggtcagacggttgacggaaaagggcag	66		
Qy	817	accttcaccgcgggttcaga---attaggcgatggcggccagctctgaaacaccagaacgcg	873		
Db	67	accttatgtgcttaatccgaatacatctggggagacggaatcgagacgcgagaatcatgaagccg	126		

```

Qy 874 ctgtttactctggaagcgtgtccagcctgaaaaacgctcaaccatggtgcg 922
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 127 atcttctgtctgagcgtggggcagcctgaaaaatgtagtattcgtg 175

RESULT 5
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgpt-fls
; US-08-232-463-14

```

	Query Match	3.4%	Score 45.8;	DB 1:	Length 7218;
	Best Local Similarity	4.9%;	Pred. NO. 0.00054;		
	Matches	11;	Conservative 135;	Mismatches 77;	Indels 0; Gaps 0;
Qy	426	gatgagcggcaaacgcatacttggccaaacctgtacaggccaacaacagtgcctcttc	485		
Dz	1048	GTCGAGGAGCTTGCATYTTTTTTT : : : : : : : : : : : : : : : :	1107		
Qy	486	cggtaactttcatcttgcggttcccccttttaaagatcatcatacaggggaagccccttc	545		
Dz	1108	VV	1167		
Qy	546	cggcaactcccttcgcggaactacttccogtcaagtacccttctcaccccattcacgcc	605		
Dz	1168	VVV	1227		
Qy	606	aacgtcccttacctaaccqcttgatttccctcttctctccacc	648		

Db 1228 YY 1270

RESULT 6

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 3.2%; Score 43.2; DB 1; Length 7218;
Best Local Similarity 5.8%; Pred. No. 0.0036;
Matches 21; Conservative 189; Mismatches 152; Indels 0; Gaps 0;

Qy 58 gggggggaacacggcttggtgcataatgcgaattctgcgttggggcaacacccatc 117
Db 1395 RRR 1326
Qy 118 gatcggaacaccattgagcaaatggctcaattattggcggaactgttaaagtcactgcta 177
Db 1325 RRR 1266
Qy 178 tcgccacaatcaggtaatgcggcaccggagccggtggaatgaccagactacaggagtt 237
Db 1265 RRR 1206
Qy 238 ggtaacgctggcgccctgaacgcgaaagacagcagcaggaaccactccgcagctgac 297
Db 1205 RRR 1146

Qy 298 agtcagaacatgctgagtgcagatgggcaaacaggggctggtgcagcccatcacgccgat 357
Db 1145 RRR 1086
Qy 358 gggcaggcgggcgagatcgcgataatcccttactgaaagcccatgctgaagcttatt 417
Db 1085 RRR 1026
Qy 418 gc 419
Db 1025 TC 1024

RESULT 7

US-08-343-443B-1/c
; Sequence 1, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain
; APPLICANT: Delattre, Olivier
; APPLICANT: Desmaze, Chantal
; APPLICANT: Melot, Thomas
; APPLICANT: Peter, Martine
; APPLICANT: Ploougastel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zucman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: AEDIT 1.0 DOS text editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,443B
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00494
; FILING DATE: 19-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/06123
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989.6121P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2371 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..1992
US-08-343-443B-1

RESULT 10

US-08-658-136-1
 ; Sequence 1, Application US/08658136
 ; Patent No. 6071717
 ; GENERAL INFORMATION:
 ; APPLICANT: KLINGER, KATHERINE W
 ; APPLICANT: LANDES, GREGORY M
 ; APPLICANT: BURN, TIMOTHY C
 ; APPLICANT: CONNORS, TIMOTHY D
 ; APPLICANT: DACKOWSKI, WILLIAM
 ; APPLICANT: GERMINO, GREGORY
 ; APPLICANT: QIAN, FENG
 ; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
 ; NUMBER OF SEQUENCES: 58
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENZYME CORPORATION
 ; STREET: ONE MOUNTAIN ROAD
 ; CITY: FRAMINGHAM
 ; STATE: MASSACHUSETTS
 ; COUNTRY: USA
 ; ZIP: 01701
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/658,136
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: LASSEN, ELIZABETH
 ; REGISTRATION NUMBER: 31,845
 ; REFERENCE/DOCKET NUMBER: GEN4-17.8
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 508-872-8400
 ; TELEFAX: 508-872-5415
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 53577 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-658-136-1

Query Match 2.7%; Score 35.8; DB 3; Length 53577;
 Best Local Similarity 57.7%; Pred. No. 2.6;
 Matches 64; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 QY 539 ccccttcgggaactcccttcgggaactactctcccgtagctctctcaccgccat 598
 ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
 Db 34644 CCCCTCC 34703
 QY 599 ccacgccaacgctccctaccctaccgctgattccctctctctccaccaca 649
 || || || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 34704 CCCTCCCTCCA 34754

RESULT 11

US-08-945-848-7
 ; Sequence 7, Application US/08945848
 ; Patent No. 5968772
 ; GENERAL INFORMATION:
 ; APPLICANT: MATSUSHIRO, Aizo
 ; TITLE OF INVENTION: PEARL PROTEIN(NACREIN) AND PROCESS FOR
 ; TITLE OF INVENTION: THE SAME
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 233 South Wacker Drive/6300 Sears Tower
 ; CITY: Chicago

STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/945,848
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Cawley, Jr, Thomas A.
 REGISTRATION NUMBER: 40,944
 REFERENCE/DOCKET NUMBER: 19036/34324
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 474-6300
 TELEFAX: (312) 474-0448
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1341 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: sig_peptide
 LOCATION: 21..51
 FEATURE:
 NAME/KEY: binding site
 LOCATION: 52..99
 FEATURE:
 NAME/KEY: repeat_region
 LOCATION: 742..969
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 724..727
 FEATURE:
 NAME/KEY: active site
 LOCATION: 1123..1266
 US-08-945-848-7

Query Match 2.6%; Score 35.2; DB 2; Length 1341;
 Best Local Similarity 47.0%; Pred. No. 0.4;
 Matches 109; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
 QY 803 atggcaaggacaaaccttcaccgcggttcagaattaggcgatggcgccagtctgaaa 862
 ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 716 ACGGCAATGGGACAAATGGCAATAACGGCTACAATGGGGACAAACGGTACAATGGTGACA 775
 QY 863 accagaaacccgtgtttacttggaagacggtgccagcctgaaaacgtcaccatggcg 922
 || || || || || || || || || || || || || || || || || || || ||
 Db 776 ACGGCAATAACAGCTACAAATGGGGACAAACGGTAACAATGGTGTCAACGGCAATAACGGCT 835
 QY 923 acgacggggcgatggtattcttcttcacgttgatgacaaatagacaatctgcacgtca 982
 || || || || || || || || || || || || || || || || || || || ||
 Db 836 ACAATGGGGACAAACGGTACAATGGGACAAACGGCAATAACGGCTACAATGGGACAACG 895
 QY 983 ccaacgtgggtgagacgcgattaccgtttaaccgcaaacagcgcgggcaaaaa 1034
 ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 896 GTACAATGGTGACAAACGGCAATAACGGTGAACGGCAATAACGGTGAACGGTGAAC 947

RESULT 12

US-08-945-848-6
 ; Sequence 6, Application US/08945848
 ; Patent No. 5968772
 ; GENERAL INFORMATION:
 ; APPLICANT: MATSUSHIRO, Aizo
 ; TITLE OF INVENTION: PEARL PROTEIN(NACREIN) AND PROCESS FOR
 ; TITLE OF INVENTION: THE SAME


```

: REFERENCE/DOCKET NUMBER: PHCR-1-8504
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: 206-682-8100
: TELEFAX: 206-225-0709
:
: INFORMATION FOR SEQ ID NO: 10:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 485 base pairs
:
: TYPE: nucleic acid
:
: STRANDEDNESS: double
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: ORIGINAL SOURCE:
:
: ORGANISM: Homo sapiens
:
: IMMEDIATE SOURCE:
:
: CLONE: 14B1
:
: FEATURE:
:
: NAME/KEY: CDS
:
: LOCATION: 3..485
:
PCT-US95-05741-10

```

	Query Match	2.5%	Score 34.2;	DB 5;	Length 485;
	Best Local Similarity	62.1%;	Pred. No. 0.44;		
	Matches 54;	Conservative 0;	Mismatches 33;	Indels 0;	Gaps 0;
QY	541	ccttcgggaactccccttccggaactactctcctccgtacgtacctttcacccccatcc	600		
DG	142	CCTTCTTCCTCCTCCTCTCTCTCTCTCTCTCTCTCCGCCAGCGCCTTCCTTGACCT	83		
QY -	601	agcgcaacgtccccctaccctcacgcgtt	627		
DG	82	CGGCCAACGTGGCCTCGCTGCCCTCTTT	56		

RESULT 15
US-08-233-788A-58/c
Sequence 58, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
; US-08-233-788A-58

Query Match          2.5%; Score 34; DB 1; Length 456;
Best Local Similarity 54.9%; Pred. No. 0.49;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 88 gcaaatctgogttggyggaacaaacccatgatcgggcaaacattgagcaaatggctcaa 147
    |||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 173 GCAAGCGCAGCGTTPAGCGGAACCGCTACTGTATAAATGCTCAACGTTGAGTCGCGGCCGGA 114
    |||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

Qy 148 ttattggcggaactgttaagtcactgtatcgccacaatcaggtaatgcggaacccgga 207
    |||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 113 CTATGTGCGCGCGCGTGTGATTACCGCGCGCGCGCCCATTTGGAACGACGCCAGCCAGA 54
    |||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

Qy 208 gc 209
    ||
Db 53 GC 52

Search completed: October 21, 2001, 11:32:22
Job time: 2700 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2001, 10:50:32 ; Search time 86.67 seconds
(without alignments)
9736.931 Million cell updates/sec

Title: US-09-596-958-1
Perfect score: 1344
Sequence: 1 atgtcaattcttaagcttaa.....acctgaagggtggatgaatga 1344

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1344	100.0	1344	20 AAX09007	Hypersensitive res
2	1344	100.0	1344	21 AAD00669	Erwinia amylovora
3	1344	100.0	1344	21 AAD014939	DNA encoding a hyp
4	155.2	11.5	1729	20 AAX24347	Hypersensitive res
5	155.2	11.5	1729	21 AAD00673	Pseudomonas syring
6	155.2	11.5	1729	21 AAA14943	DNA encoding a hyp
7	94.8	7.1	591	19 AAV69879	Bacillus sp strain
8	94.8	7.1	591	19 AAV59478	Bacillus sp. pecti
9	74.2	5.5	666	20 AAX89484	Bacillus lichenifo
10	46.6	3.5	732	21 AAF12333	Aspergillus oryzae
11	42.8	3.2	1166	20 AA219440	M. tuberculosis an

12	42.8	3.2	1166	20	AA219228	M. tuberculosis re
13	38.2	2.8	1925	20	AAX90924	Epstein Barr Virus
14	37.6	2.8	4289	19	AAV62147	HSV-2 strain SB5 C
15	37.4	2.8	2371	14	AAQ50643	Human Ews gene clo
16	37.4	2.8	3309	14	AAQ50646	Human Ews gene (ge
17	37	2.8	117213	19	AAV62176	HSV-2 strain SB5 C
18	36.4	2.7	1320	22	AAF72217	Corynebacterium gl
19	36.2	2.7	400	19	AAV27417	Streptococcus pneu
20	36.2	2.7	3510	19	AAV52186	Streptococcus pneu
21	35.8	2.7	1370	21	AAI13346	Aspergillus oryzae
22	35.8	2.7	3453	21	AAV79709	Eucalyptus grandis
23	35.8	2.7	53526	19	AAI94101	Human PKD1 gene.
24	35.8	2.7	53577	17	AAI18551	Human polycystic k
25	35.8	2.7	53577	19	AAI94108	Human PKD1 locus b
26	35.8	2.7	114955	20	AAX53491	Human adenosine Al
27	35.4	2.6	335	21	AAI44637	Human secreted exp
28	35.2	2.6	2363	17	AAI58313	Oyster pearl prote
29	34.8	2.6	1335	21	AAI77918	Human cancer assoc
30	34.4	2.6	596	21	AAI14025	Aspergillus oryzae
31	34.4	2.6	6032	19	AAV30268	Plasmid pMT1802 en
32	34.4	2.6	6032	19	AAV12368	Trichoderma harzia
33	34.2	2.5	485	16	AAI05516	Human neurogenic d
34	34.2	2.5	1337	20	AAI17263	Human gene express
35	34.2	2.5	114955	20	AAX53491	Human adenosine Al
36	34	2.5	456	15	AAQ87467	Agfa sequence. Sa
37	34	2.5	456	18	AAI74142	Salmonella enterit
38	34	2.5	456	21	AAI64617	Salmonella enterit
39	34	2.5	456	21	AAI64622	Agfa::PT3#1 DNA se
40	34	2.5	456	21	AAI64623	Agfa::PT3#2 DNA se
41	34	2.5	456	21	AAI64626	Agfa::PT3#5 DNA se
42	34	2.5	456	21	AAI64627	Agfa::PT3#6 DNA se
43	34	2.5	456	21	AAI64628	Agfa::PT3#7 DNA se
44	34	2.5	456	21	AAI64630	Agfa::PT3#9 DNA se
45	34	2.5	456	21	AAI64631	Agfa::PT3#10 DNA s

ALIGNMENTS

RESULT 1
AAX09007
ID AAX09007 standard; DNA; 1344 BP.
XX AC AAX09007;
XX DT 14-JUN-1999 (first entry)
XX DE Hypersensitive response eliciting protein coding sequence (hrpW).
XX KW Hypersensitive response elicitor protein; hypersensitive response;
KW hrpW; pathogen; infection; crop protection; disease resistance;
KW pest resistance; transgenic plant; colouration; maturation; ss.
XX OS Erwinia amylovora.
XX FH Key Location/Qualifiers
FT CDS 1..1344
FT /*tag= a
FT /product= Hypersensitive_response_elicitor_protein
XX WO9907208-A1.
XX PD 18-FEB-1999.
XX PF 27-JUL-1998; 98WO-US15547.
XX PR 06-AUG-1997; 97US-0055108.
XX PA (CORR) CORNELL RES FOUND INC.
XX PI Beer SV, Kim JF;
XX WPI; 1999-167126/14.
DR

DR P-PSDB; AAW96260.
XX New Erwinia amylovora hypersensitive response eliciting gene and
PT protein - useful for providing transgenic plants and seeds with
PT enhanced growth, and insect and disease resistance
XX
PS Claim 1: Page 49; 54pp; English.
XX
CC The hypersensitive response eliciting protein (hrp) or polypeptide
CC is produced as part of an active defense by plants against
CC incompatible pathogen infections. The hypersensitive response is a
CC rapid localised necrosis. The hrp protein and gene when used in
CC nucleotide constructs are useful for providing disease resistance to
CC plants, insect control to plants, and enhancing plant growth
CC (enhancing fruit size and earlier colouration and maturation), by
CC direct application of the protein to plants, or by producing
CC transgenic plants or seeds using the hrp gene.
XX
SQ Sequence 1344 BP; 343 A; 372 C; 357 G; 272 T; 0 other;

Query Match 100.0%; Score 1344; DB 20; Length 1344;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtcaattcttacgcttaacaacaataactcgtctcctcgcgggtgtgttccagtcgcgg 60
DB 1 atgtcaattcttacgcttaacaacaataactcgtctcctcgcgggtgtgttccagtcgcgg 60

QY - 61 ggggacaacggcgttggtgataatgaaattctcgttggggcaacaacccatcgat 120
DB 61 ggggacaacggcgttggtgataatgaaattctcgttggggcaacaacccatcgat 120

QY 121 caggcaaacattgacaaatggctcaattattgctgggaactgttaaagtcactgctatcg 180
DB 121 caggcaaacattgacaaatggctcaattattgctgggaactgttaaagtcactgctatcg 180

QY 181 ccacaatacaggtaattcggcaacccggagccggtggtgcaatgaccagactacaggagtgg 240
DB 181 ccacaatacaggtaattcggcaacccggagccggtggtgcaatgaccagactacaggagtgg 240

QY 241 aacgctggcggcctgaacgagaaagacacagcaggaacacactccgactcgcagct 300
DB 241 aacgctggcggcctgaacgagaaagacacagcaggaacacactccgactcgcagct 300

QY 301 cagaacatgctgagtgagatgggcaacaacggcgtgagtcagccatcacgcccgatggc 360
DB 301 cagaacatgctgagtgagatgggcaacaacggcgtgagtcagccatcacgcccgatggc 360

QY 361 caggcggcggcggcagatcggcgataatcttttactgaaagccaatgctgaagcttattgca 420
DB 361 caggcggcggcggcagatcggcgataatcttttactgaaagccaatgctgaagcttattgca 420

QY 421 cgcgatgagcggcccaacgcatcagtttgccaaacctggtacgggcaacaacagtgc 480
DB 421 cgcgatgagcggcccaacgcatcagtttgccaaacctggtacgggcaacaacagtgc 480

QY 481 tcttcggctacttcttcactcgtgggttcccttttaacgatctatcaggggggaaggcc 540
DB 481 tcttcggctacttcttcactcgtgggttcccttttaacgatctatcaggggggaaggcc 540

QY 541 ccttcggcaactcccttcggcaactactctcccgtaagtaacctctcaaccccatcc 600
DB 541 ccttcggcaactcccttcggcaactactctcccgtaagtaacctctcaaccccatcc 600

QY 601 agcgaacatccctaccctcacgcttgattctctcttctcccaacaaagcagccgg 660
DB 601 agcgaacatccctaccctcacgcttgattctctcttctcccaacaaagcagccgg 660

QY 661 ggcagcacccggtaaccgatcatcctgacctgttggtagcgcgggcatcgggccgga 720
DB 661 ggcagcacccggtaaccgatcatcctgacctgttggtagcgcgggcatcgggccgga 720

QY 721 aattcgggtgcttccaccgagccggtgctaaatcagacggtgctgcatgacaccattacc 780
DB 721 aattcgggtgcttccaccgagccggtgctaaatcagacggtgctgcatgacaccattacc 780

QY 781 gtgaaagcgggtcaggtgttggatggcaagagacaaacccctaccgccggttcagaaatta 840
DB 781 gtgaaagcgggtcaggtgttggatggcaagagacaaacccctaccgccggttcagaaatta 840

QY 841 ggcgatggcggccagctctgaaacacagaaacgcgtgtttatattctggaagcgggtgccag 900
DB 841 ggcgatggcggccagctctgaaacacagaaacgcgtgtttatattctggaagcgggtgccag 900

QY 901 ctgaaaaacgtcaccatggcgacgacggggcggtatggtattcatctttacggtgatgcc 960
DB 901 ctgaaaaacgtcaccatggcgacgacggggcggtatggtattcatctttacggtgatgcc 960

QY 961 aaaaatagacaattctgcacgtcaccacacgtgggtgagagcgcgattaccgttaagccaaac 1020
DB 961 aaaaatagacaattctgcacgtcaccacacgtgggtgagagcgcgattaccgttaagccaaac 1020

QY 1021 agcgcgggcacaaaaatccacgttgaaatcaactaactacagttccttcgagcacgcctctgac 1080
DB 1021 agcgcgggcacaaaaatccacgttgaaatcaactaactacagttccttcgagcacgcctctgac 1080

QY 1081 aagatcctgcagctgaatgccgataactaacctgagcgttgacaacgtgaaggccaaagac 1140
DB 1081 aagatcctgcagctgaatgccgataactaacctgagcgttgacaacgtgaaggccaaagac 1140

QY 1141 ttgtgtaatttgcacgtaactaacgcggcgtcaacaggttaactgggatctgaatctgagc 1200
DB 1141 ttgtgtaatttgcacgtaactaacgcggcgtcaacaggttaactgggatctgaatctgagc 1200

QY 1201 catatcagcgcgagaacggttaagtctctcgttctgaaagcgatagcagggggtctaaac 1260
DB 1201 catatcagcgcgagaacggttaagtctctcgttctgaaagcgatagcagggggtctaaac 1260

QY 1261 gtcaatcaccagtgatctcactcgtggtgtgttgaaacacactacaaaagtgcgagtcc 1320
DB 1261 gtcaatcaccagtgatctcactcgtggtgtgttgaaacacactacaaaagtgcgagtcc 1320

QY 1321 gccaaacctgaagtggtgctgaatga 1344
DB 1321 gccaaacctgaagtggtgctgaatga 1344

RESULT 2
AAD00669
ID AAD00669 standard; DNA; 1344 BP.
XX
AC AAD00669;
XX
DT 08-SEP-2000 (first entry)
XX
DE Erwinia amylovora hypersensitive response elicitor encoding DNA #2.
XX
KW Hypersensitive response elicitor; environmental stress resistance;
KW plant; ds.
XX
OS Erwinia amylovora.
XX
FH Key Location/Qualifiers
FT 1..1344
FT /*tag= a
FT /product= "Hypersensitive response elicitor"
XX
XX WO200028055-A2.
PN
XX 18-MAY-2000.
PD
XX 04-NOV-1999; 99WO-US26039.
XX
XX 05-NOV-1998; 98US-0107243.
XX

PA (EDEN-) EDEN BIOSCIENCE CORP.

XX Wei Z, Schading RL;

XX WPI; 2000-376566/32.

DR P-PSDB; AAY1094.

XX Application of a hypersensitive response elicitor protein to plants to

PT Impart stress resistance -

XX Disclosure; Page 10; 84pp; English.

XX The patent discloses a method to impart stress resistance to plants by

CC applying a hypersensitive response elicitor in a non-infectious form to

CC a plant or seed. The present sequence is a DNA encoding hypersensitive

CC response elicitor protein from Erwinia amylovora. The protein is

CC heat stable, protease sensitive and suppressed by inhibitors of plant

CC metabolism. The present sequence is used to transform transgenic plant

CC or plant seeds to impart stress resistance.

XX

SQ Sequence 1344 BP; 343 A; 372 C; 357 G; 272 T; 0 other;

Query Match 100.0%; Score 1344; DB 21; Length 1344;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	atgtcaatttcacgttaacaaataacacctctcctcgcgggtctgttccagtcggg	60
Db	1	atgtcaatttcacgttaacaaataacacctctcctcgcgggtctgttccagtcggg	60
Qy	61	ggggacaacgggtggtggtcataatgcaaatctcgttggggcaacacccatcgat	120
Db	61	ggggacaacgggtggtggtcataatgcaaatctcgttggggcaacacccatcgat	120
Qy	121	cggcaaacattgagcaaaatggctcaattattggcggaactgttaaagtcaactgctatcg	180
Db	121	cggcaaacattgagcaaaatggctcaattattggcggaactgttaaagtcaactgctatcg	180
Qy	181	ccacaatcaggtaatcgggcaacggagcgggtggcgaatgaccagactacaggaattggt	240
Db	181	ccacaatcaggtaatcgggcaacggagcgggtggcgaatgaccagactacaggaattggt	240
Qy	241	aacgctggcgccctgaacgagcaaaaaggcacacgacgaaacacactccgacgtctgacagt	300
Db	241	aacgctggcgccctgaacgagcaaaaaggcacacgacgaaacacactccgacgtctgacagt	300
Qy	301	cagaacatgctgagtgagatgggcaacaacgggctggtgatcaggccatcacgcccagtgcc	360
Db	301	cagaacatgctgagtgagatgggcaacaacgggctggtgatcaggccatcacgcccagtgcc	360
Qy	361	caggcgccggcgagatcgccgataatccctttactgaagccatgctgaagcttatgca	420
Db	361	caggcgccggcgagatcgccgataatccctttactgaagccatgctgaagcttatgca	420
Qy	421	cgcgatggagccgcaaaagcagtcagtttggccaacctggtacggcgcaacaacagtgcc	480
Db	421	cgcgatggagccgcaaaagcagtcagtttggccaacctggtacggcgcaacaacagtgcc	480
Qy	481	tcctccggtaacttcttcacatctggcgggttcctcttttaacgatctatcagggggaaagcc	540
Db	481	tcctccggtaacttcttcacatctggcgggttcctcttttaacgatctatcagggggaaagcc	540
Qy	541	cttccggcaactcccttcgcgaactactctccctcagtcactctcaccctatcc	600
Db	541	cttccggcaactcccttcgcgaactactctccctcagtcactctcaccctatcc	600
Qy	601	acgcaacgctccctacactaccgcttgatttcccttcttccaccacaaagcagccggg	660
Db	601	acgcaacgctccctacactaccgcttgatttcccttcttccaccacaaagcagccggg	660
Qy	661	ggcagcagccgggtaacagatcatactgacacctgttgtagcgcggtatcggggcccga	720
Db	661	ggcagcagccgggtaacagatcatactgacacctgttgtagcgcggtatcggggcccga	720

Db	661	ggcagcagccgggtaacagatcatactgacacctgttgtagcgcggtatcggggcccga	720
Qy	721	aattcgtgaccttcacacgcccgcgtaataatcagacgggtgctgcatgacacattacc	780
Db	721	aattcgtgaccttcacacgcccgcgtaataatcagacgggtgctgcatgacacattacc	780
Qy	781	gtgaagcgggtcaggtgtttgatggcaaaagacaaaccttcaccgcggttcagaatta	840
Db	781	gtgaagcgggtcaggtgtttgatggcaaaagacaaaccttcaccgcggttcagaatta	840
Qy	841	ggcgatggcgccagctctgaaacacagaaacccgctgtttatactggaagcgggtgcacgc	900
Db	841	ggcgatggcgccagctctgaaacacagaaacccgctgtttatactggaagcgggtgcacgc	900
Qy	901	ctgaaaaacgctcacatggcgacgacggcggtggtattcttcttccggtgagtcgc	960
Db	901	ctgaaaaacgctcacatggcgacgacggcggtggtattcttcttccggtgagtcgc	960
Qy	961	aaaatagacaatctgcgctcaccacacgctgggtgagacgcgattaccgtttaagccaaac	1020
Db	961	aaaatagacaatctgcgctcaccacacgctgggtgagacgcgattaccgtttaagccaaac	1020
Qy	1021	agcgcgggcaaaaaatcccccgcttgaaatacactaacagttctctcggagcagcctctgac	1080
Db	1021	agcgcgggcaaaaaatcccccgcttgaaatacactaacagttctctcggagcagcctctgac	1080
Qy	1081	aagatctctgcagctgaatgcgatacactaacctgagcgttgacaacgtaagggccaaagac	1140
Db	1081	aagatctctgcagctgaatgcgatacactaacctgagcgttgacaacgtaagggccaaagac	1140
Qy	1141	tttgactctttgtacgcactaacgcggttcaacagggtaactggatctgaatctgagc	1200
Db	1141	tttgactctttgtacgcactaacgcggttcaacagggtaactggatctgaatctgagc	1200
Qy	1201	catacgcgcgacgaagcggtaagtctcgttctttaaagcagatagcaggggctaaac	1260
Db	1201	catacgcgcgacgaagcggtaagtctcgttctttaaagcagatagcaggggctaaac	1260
Qy	1261	gtcaataccagtgatctcactgggtgagttgaaacacactacaagtcgcgagtcgc	1320
Db	1261	gtcaataccagtgatctcactgggtgagttgaaacacactacaagtcgcgagtcgc	1320
Qy	1321	gccaaacctgaaggtggtcgaatga 1344	
Db	1321	gccaaacctgaaggtggtcgaatga 1344	

RESULT 3

AAAL4939

ID AAAL4939 standard; DNA; 1344 BP.

XX AAAL4939;

AC AAAL4939;

XX 08-AUG-2000 (first entry)

DE DNA encoding a hypersensitive response elicitor protein.

XX Hypersensitive response; insect control; disease resistance;

XX Hypersensitive response elicitor; plant growth; vegetable; crop;

XX ornamental plant; ss.

XX Erwinia amylovora.

XX Key Location/Qualifiers

FT CDS 1..1344

FT /*tag= a

FT /product= "hypersensitive response elicitor protein"

PN WO200020452-A2.

XX 13-APR-2000.

XX 05-OCT-1999; 99WO-US23181.

XX Key Location/Qualifiers
 FT CDS 419..1693
 FT /*tag= a
 XX
 PN WO9907207-A1.
 XX
 PD 18-FEB-1999.
 XX
 XX 24-JUL-1998; 98WO-US15501.
 XX
 PF 06-AUG-1997; 97US-0055107.
 XX
 PR (CORR) CORNELL RES FOUND INC.
 XX
 PA Alfano JR, Charkowski A, Collmer A;
 XX
 XX WPI; 1999-167125/14.
 DR P-PSDB; AAW97851.
 XX
 XX New hypersensitive response eliciting (dsPE) gene and protein -
 PT useful for providing transgenic plants and seeds with enhanced
 PT growth, and insect and disease resistance
 XX
 PS Claim 1: Page 37-38; 56pp; English.
 XX
 CC This the DNA sequence of the dsPE gene of *Pseudomonas syringae*
 CC pv. tomato DC3000 that codes for a 42.9 kDa hypersensitive response
 CC elicitor (HRE) polypeptide (see AAW97851), termed hrpW. To identify
 CC any HRE-like genes in the *P. syringae* pv. tomato DS3000 DNA flanking
 CC hrpR, cosmid pCP2357, which contains this region in vector pCP247,
 CC was isolated. A series of subclones in pML123 were constructed and
 CC screened for 2 potential HR phenotypes: (i) the ability to promote
 CC tobacco HRE activity in *Pseudomonas fluorescens* cells carrying
 CC pCP2274, a delhrpZ *phrII* derivative, and (ii) interference with
 CC the HRE activity of *P. fluorescens* cells carrying wild-type *phrII*.
 CC No subclones had the first phenotype, but one, pCP2373, had the
 CC second. Transcriptional unit V of this subclone contained the
 CC 1275 bp open reading frame encoding HrpW. The HrpW protein or
 CC isolated DNA molecule can be used to impart disease resistance to
 CC plants, to enhance plant growth and/or to control insects on
 CC plants. This is achieved by applying the HRE protein in a
 CC non-infectious form to plants or plant seeds. Alternatively,
 CC transgenic plants or plant seeds transformed with DNA encoding the
 CC HRE can be provided.
 XX
 SQ Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 other;

Query Match 11.5%; Score 155.2; DB 20; Length 1729;
 Best Local Similarity 57.1%; Pred. No. 9.6e-37;
 Matches 330; Conservative 0; Mismatches 233; Indels 15; Gaps 2;

QY 745 ggcgctaatacagcgtgtctgcatgacacattaccgtgaaagcggttcaggtttgat 804
 DB 1079 gccggaagatcaatgtgtgaaagacaccatcaagtcgctgctgcaagtctttgac 1138
 QY 805 gcaaaaggacaaccttcaccgcttcagaattaggcgtgcgcagctctgaaac 864
 DB 1139 ggcacggcgcaaccttcactgcgcacaattctatggttaacggagaccaggcgcaaat 1198
 QY 865 cagaaacgcgtgtttactggaagcgtgtccagctgaaacacgtccacctggcgac 924
 DB 1199 cagaagccatgtctgagctggtgagcgctacgtcagtgaagaattgaacctgggtgag 1258
 QY 925 gacggggcggtatggttattcatctttacg-----gtgatccaaaatagacaat 972
 DB 1259 aacgaggtcgatggtatccacgtgaaagcacaacacgtcaggaagtccattgacaac 1318
 QY 973 ctgcagctaccacagctgggtgagcgagcttaccgttaagcacaacagcgcggaacaa 1032
 DB 1319 gtgcgtatcccaagacgtcgtgtgaaacacctgattcacgttcaaaagcgagggcgagcg 1378

QY 1033 aaatccacggttgaataactactaatacagttcctctgagcacgcctctgacaagactctgcag 1092
 DB 1379 gtcaataatctgaacatcaagaacagcagtgccaaggttcagacagacaaggtgttcag 1438
 QY 1093 ctgaatgccgataactaacctgagcgttgacaacgctgaagccaaagactttgtactttt 1152
 DB 1439 ctcaacgccaaactcacttgaaatcgacaacttcaagccgacgatttctcgacgatg 1498
 QY 1153 gtacgcactaacgcgcgtcaacag---ggtaaactgggatctgaatctgagccatatacgc 1209
 DB 1499 gttcgaccaaacggtggaagcagtttgatgacatgacatcagctgaacgcgcatcgaa 1558
 QY 1210 gcagaagacggttaagttctctgctttaaagcgatagcaggggttaaacgctcaatacc 1369
 DB 1559 gctaaccacggcaagttcgccctggtgaaagcgacagtgacgatctgaagctggcaacg 1618
 QY 1270 agtgatatctcactgggtgatgtgtgaaacccactacaa 1307
 DB 1619 ggcaacatgcctgacccgacgtcaaacacgcctacga 1656

RESULT 5
 AAD00673
 ID AAD00673 standard; DNA; 1729 BP.
 XX
 AC AAD00673;
 XX
 DT 08-SEP-2000 (first entry)
 XX
 DE *Pseudomonas syringae* hypersensitive response elicitor encoding dsPE gene.
 XX
 KW Hypersensitive response elicitor; environmental stress resistance;
 KW plant; dsPE gene; ds.
 XX
 OS *Pseudomonas syringae*.
 XX
 FH Key Location/Qualifiers
 FT CDS 419..1693
 FT /*tag= a
 FT /product= "Hypersensitive response elicitor"
 XX
 PN WO200028055-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 04-NOV-1999; 99WO-US26039.
 XX
 PR 05-NOV-1998; 98US-0107243.
 XX
 PA (EDEN-) EDEN BIOSCIENCE CORP.
 XX
 XX Wei Z, Schading RL;
 PT
 XX WPI; 2000-376566/32.
 DR P-PSDB; AAY71098.
 XX
 PT Application of a hypersensitive response elicitor protein to plants to
 PT impart stress resistance
 XX
 PS Disclosure; Page 24-25; 84pp; English.
 XX
 CC The patent discloses a method to impart stress resistance to plants by
 CC applying a hypersensitive response elicitor in a non-infectious form to
 CC a plant or seed. The present sequence is dsPE gene encoding
 CC hypersensitive response elicitor protein from *Pseudomonas syringae*.
 CC The present sequence is used to transform transgenic plant or plant
 CC seeds to impart stress resistance.
 XX
 SQ Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 other;

Query Match 11.5%; Score 155.2; DB 21; Length 1729;
 Best Local Similarity 57.1%; Pred. No. 9.6e-37;

Matches 330; Conservative 0; Mismatches 233; Indels 15; Gaps 2;

```
Qy 745 ggcgcataatcagcgtctgcatcacaccattaccgtgaaagcgggtcaggtgtttat 804
Db 1079 gccggcagatcaatgctgtgaaagaccaccatcaaggctggcgctggaagctcttgac 1138
Qy 805 ggcaaaaggacaaaccttcacgcgggttcagaatagcgatggcgccagctctgaaaac 864
Db 1139 ggcacggcgcaaccttcacgtccgacaaatctatgggtaacggagaccaggcgcaaac 1198
Qy 865 cagaaacgcgtgtttactctggaagacagcgtgccaagcctgaaacacgtccaccatggcgac 924
Db 1199 cagaagccatgttcgagctgctgaagcgtactagcttgaagaatgtgaacctgggtgag 1258
Qy 925 gacggggcgaggtgattcatcttaacg-----gtgatgccaaatagacaaat 972
Db 1259 aacgaggtcgatggcatccacgtgaaagcccaaaacgctcaggaagtcaccattgacaaac 1318
Qy 973 ctgcacgtcaccaacgctgggtgagcgcgtattaccgtttaagccaaacagcggcgcaaa 1032
Db 1319 gtgcacgtcccaagacgtcgggtgaagacctgattacggttcaaaagcgagggcgacgag 1378
Qy 1033 aaatccacgttgaaatcacataacagttccttcgagacgcgcctctgacaagatcctgcag 1092
Db 1379 gtcaataatctgaacatcaagaacagcagtgccaaaggtgcagacagaaggtgtccag 1438
Qy 1093 ctgaatgcgcataactaacctgacgcttgacacgtgaaaggccaaagactttgttactttt 1152
Db 1439 ctcaacgcacaacactcaacttgaaatcgcaaaactcaaggcgcagatttcggcacgatg 1498
Qy 1153 gtacgcataacgcgggttcacag---ggttaactgggattctgaatctgagccataatcagc 1209
Db 1499 gttcgcaccaacggctggcaagcagtttgcagacagcagctgagctgagcggcatcgaa 1558
Qy - 1210 gcagaagcggtaagtctctgtctgtaaaagcagatagcggggcgtaaagctcaatacc 1269
Db 1559 gctaaccacggcaagttcgcctctgtgaaaagcgacagtgcagatctgaagctggcaacg 1618
Qy 1270 agtgatatctcactgggtgatgttgaaacacactacaa 1307
Db 1619 ggcaacatcgccatgaccgacgttcaaacacgcctacga 1656
```

```
RESULT 6
AAAL14943
ID AAAL14943 standard; DNA; 1729 BP.
XX
AC AAAL14943;
XX
DT 08-AUG-2000 (first entry)
XX
DE DNA encoding a hypersensitive response elicitor protein.
XX
KW Hypersensitive response; Insect control; disease resistance;
KW hypersensitive response elicitor; plant growth; vegetable; crop;
KW ornamental plant; ss.
XX
OS Pseudomonas syringae.
XX
FH Key Location/Qualifiers
FT CDS 419..1693
FT /*tag= a
FT /product= "hypersensitive response elicitor protein"
XX
PN W0200020452-A2.
XX
PD 13-APR-2000.
XX
PF 05-OCT-1999; 99WO-US23181.
XX
PR 05-OCT-1998; 98US-0103050.
XX
FA (EDEN-) EDEN BIOSCIENCE CORP.
```

```
XX Wei Z, Fan H, Niggemeyer JL;
PI
XX WPT: 2000-303745/26.
DR P-FSDB: AAY84859.
XX
PT Hypersensitive response elicitor polypeptides useful for imparting
PT enhanced growth, disease resistance and insect resistance to plants,
PT especially vegetables and ornamental flowers -
XX
PS Disclosure: Page 25-26; 100pp; English.
XX
CC The present sequence encodes a hypersensitive response elicitor
CC polypeptide. The specification describes hypersensitive response
CC elicitor polypeptide fragments, which do not elicit a hypersensitive
CC response. Instead, the proteins impart disease resistance to plants,
CC enhance plant growth, and/or control insects. The polypeptide
CC fragments may be used to these properties to plants. The plants which
CC may be treated in this way include vegetables, crops and ornamental
CC plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,
CC peanut, corn, potato, sweet potato, bean, pea, chickory, lettuce,
CC endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,
CC broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,
CC carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,
CC citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,
CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia,
CC petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.
XX
SQ Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 Other;
```

```
Query Match 11.5%; Score 155.2; DB 21; Length 1729;
Best Local Similarity 57.1%; Pred. No. 9.6e-37;
Matches 330; Conservative 0; Mismatches 233; Indels 15; Gaps 2;
Qy 745 ggcgcataatcagcgtctgcatcacaccattaccgtgaaagcgggtcaggtgtttat 804
Db 1079 gccggcagatcaatgctgtgaaagaccaccatcaaggctggcgctggaagctcttgac 1138
Qy 805 ggcaaaaggacaaaccttcacgcgggttcagaatagcgatggcgccagctctgaaaac 864
Db 1139 ggcacggcgcaaccttcacgtccgacaaatctatggttaacggagaccaggcgcaaac 1198
Qy 865 cagaaacgcgtgtttactctggaagacgctgccaagcctgaaacacgtccaccatggcgac 924
Db 1199 cagaagccatgttcgagctggtgaagcgcgtactgtaagaatgtgaacctgggtgag 1258
Qy 925 gacggggcgaggtgtattcatctttacg-----gtgatgccaaatagacaaat 972
Db 1259 aacgaggtcgatggcatccacgtgaaagcccaaaacgctcaggaagtcaccattgacaaac 1318
Qy 973 ctgcacgtcaccaacgctgggtgagcgcgtattaccgtttaagccaaacagcggcgcaaa 1032
Db 1319 gtgcacgtcccaagacgtcgggtgaagacctgtatcaggctcaaaaggcgaggggcgacg 1378
Qy 1033 aaatccacgttgaaatcacataacagttccttcgagacgcgcctctgacaagatccctgcag 1092
Db 1379 gtcataatctgaacatcaagaacagcagtgccaaaggtgcagacgacaaggtttgcceag 1438
Qy 1093 ctgaatgcgcataactaacctgagcgttgacaacgctgaaggccaaagactttggtactttt 1152
Db 1439 ctcaacgcacaacactcaacttgaaatcgcaaaactcaaggcgacgatttcggcgacgatg 1498
Qy 1153 gtacgcataacgcgggttcacag---ggttaactgggattctgaatctgagccataatcagc 1209
Db 1499 gttcgcaccaacggctggcaagcagtttgcagacagcagctgagctgagcggcatcgaa 1558
Qy 1210 gcagaagcggtaagtctctgtctgtaaaagcagatagcggggcgtaaagctcaatacc 1269
Db 1559 gctaaccacggcaagttcgcctctgtgaaaagcgacagtgcagatctgaagctggcaacg 1618
Qy 1270 agtgatatctcactgggtgatgttgaaacacactacaa 1307
```

Db 1619 ggcaacatcgccatgaccgacgtcaaacacgcctaaga 1656

RESULT 7

ID AAV69879 standard; DNA; 591 BP.

XX AAV69879;

XX 28-JAN-1999 (first entry)

XX *Bacillus* sp strain KSM-P15 pectic acid lyase encoding DNA.

KW *Bacillus* sp. strain KSM-P15; pectic acid lyase; protopectinase;
KW protopectin; polygalacturonic acid; detergent; cotton; surfactant;
KW cellulase; protease; bleaching agent; ds.

XX *Bacillus* sp.

XX Key Location/Qualifiers
FH 1..591
CDS /*tag= a

XX WO9845393-A2.

XX 15-OCT-1998.

XX 08-APR-1998; 98WO-JP01613.

XX 08-SEP-1997; 97JP-0242736.

XX 09-APR-1997; 97JP-0091142.

XX (KAOS) KAO CORP.

XX Hatada Y, Ito S, Kasai M, Kobayashi T, Koike K;

XX Shikata S, Suzumatsu A, Tsumadori M, Wada Y;

XX WPI; 1998-568339/48.

XX Detergent composition containing protopectinase active at alkaline
PT pH - on protopectin and polygalacturonic acid, provides better
PT removal of muddy soil

XX Example; Page 71-72; 80pp; English.

XX The present invention describes a detergent composition which contains
CC a protopectinase having an optimum pH 7 or higher against protopectin
CC and polygalacturonic acid substrates. Inclusion of protopectinase
CC gives a composition that provides better removal of muddy soil,
CC particularly from socks. The present sequence encodes pectic acid lyase
CC from *Bacillus* sp. strain KSM-P15, which is used in an example from
CC the present invention. Pectic acid lyase exhibits protopectinase
CC activity.

XX Sequence 591 BP; 160 A; 134 C; 179 G; 118 T; 0 other;

Query Match 7.1%; Score 94.8; DB 19; Length 591;

Best Local Similarity 52.8%; Pred. No. 1.1e-18;

Matches 263; Conservative 0; Mismatches 217; Indels 18; Gaps 2;

QY 754 cagacggtgctgcatacacaccattaccgtgaaagcgggtcaggtgttgatgcaaga 813

Db 4 ccgaggtcgcttataaagcattcgtgctgcggtcagacgcttgacggaagaagg 63

QY 814 caaaccttaccgcgggttcaga---attagcgatggcgccagctgaaaccagaaa 870

Db 64 cagacctatgtgctaatacgaatacatatggggagcgatcgcagcggaagatcagaag 123

QY 871 ccgctgtttatactgaagacggtgccagcctgaaacgtcaaccatggcgacgaggg 930

Db 124 ccgattcttctgagggtggggcgaagcctgaaacgtgaaatgattgattggcctctgcc 183

QY 931 gcggatggtattcatctttacggtgatgccaaaatagacaatctgcagctcaccacggtg 990
Db 184 gctgacgggggtgctacgtctacgggattgtacgattacaatgtcatctggggagatgtt 243
QY 991 ggtgagacgcgattaccgtttaagccaaacagcgcggaacaaatcccccgcttgaatc 1050
Db 244 ggtgaggtgctgacgcttaaatcgtccggaacg-----gtgaacatc 288
QY 1051 actaacagttccttcagacgcgctctgacaagatctgcagctgcagtgatgcgatactaac 1110
Db 289 tcggcggggcagcctacaagcggtgacaaggtgttccaatcaatgacgagcgggagc 348
QY 1111 ctgagcgttgacaacgagcgaagccttggtactttgtactttgtacgcactaacgcggt 1170
Db 349 atcaattcgtacttcaggcgcatcgacatcggggaagctggttcggcgagacggggc 408
QY 1171 caacgggttaactggatctgaatctgagccatatcagcgacagaagcggtaagttctcg 1230
Db 409 accactacaagtggtgatgaacgtggaaaactgcaacatttccagagtgaaggtgcg 468
QY 1231 ttggttaaaagcgatagc 1248
Db 469 atcctgagaacgggacagc 486

RESULT 8

AAV59478

ID AAV59478 standard; DNA; 591 BP.

XX AAV59478;

XX 07-JAN-1999 (first entry)

XX *Bacillus* sp. pectic acid lyase encoding DNA.

XX *Bacillus* sp. KSM-P15; pectic acid lyase; pectinic acid lyase; pectin;
XX detergent; food-processing; fibre-processing agent; cell wall; ds.

XX *Bacillus* sp.

XX Key Location/Qualifiers
FH 1..591
CDS /*tag= a
FT /note= "no stop codon given"

XX EP870834-A1.

XX 14-OCT-1998.

XX 09-APR-1998; 98EP-0106586.

XX 08-SEP-1997; 97JP-0242735.

XX 09-APR-1997; 97JP-0091142.

XX (KAOS) KAO CORP.

XX Hatada Y, Ito S, Kobayashi T, Koike K, Suzumatsu A;

XX Yoshimatsu T;

XX WPI; 1998-523159/45.

XX P-PSDB; AAW77412.

XX New *Bacillus* pectic acid lyase - useful as a detergent component, a
XX food-processing agent and a fibre-processing agent

XX Claim 1; Page 16-17; 29pp; English.

XX The present sequence encodes a pectic acid lyase isolated from

CC microorganism *Bacillus* sp. KSM-P15. The pectic acid lyase has high
CC pectic acid lyase activity which degrades pectin in plant cell walls
CC and fibre in vegetables, and so is useful as a component of detergents,
CC a food-processing agent, or a fibre-processing agent. The pectic acid

Query Match	3.2%	Score 42.8	DB 20	Length 1166
Best Local Similarity	28.6%	Pred. No. 0.0086		
Matches	68	Conservative	55	Mismatches 115; Indels 0; Gaps
Qy	448	tttggccaacctggtacgggcaacaacagtcgctctctccggctactctctcatctgcgcgt	507	
		: :		: : : :
Db	563	tytnnccamcscctctmtcmaactckccgggycncmyctctckcaynmaackkky	622	
Qy	508	tccctcttttaacgatctatcaggggggaagcccctctcgggaactcccccttcgggcaac	567	
		: : : : : : : : : : : : : : : : : : :		
Db	623	cywonwymyckckagwyknmctccwactctmynttctctctcnkccmkacackntct	682	
Qy	568	tactctccctcagctactcttcacccccatccacgcaacgctcccctactccacgcgt	627	
		: : : : : : : : : :		
Db	683	cwsccccccacakayncyayawmtctcmctckacscocyynncymnmcwmtcwtclw	742	
Qy	628	gatttccctctctctcccaacaagcagccggggggcagcagccgggaaccgatcatc	685	
		: : : : : : : : : : : :		
Db	743	nakcancntctctctctmymymnackcwcnnctcncksgaccyctctactkmkckcm	800	

RESULT	12
AAZ19228	
ID	AAZ19228 standard; cDNA; 1166 BP.
XX	
AC	AAZ19228;
XX	
DT	05-NOV-1999 (first entry)
XX	
DE	- M. tuberculosis recombinant antigen cDNA encoding 3' Erdsn-8.
XX	
DE	Antigen; diagnosis; detection; infection; antibody; immunisation;
KW	vaccine; immunity; ss.
XX	
OS	- Mycobacterium tuberculosis:
XX	
PN	W09942118-A2.
XX	
PD	26-AUG-1999.
XX	
PF	17-FEB-1999; 99WO-US03265.
XX	
PR	05-MAY-1998; 98US-0072596.
XX	
PR	18-FEB-1998; 98US-0024753.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI	Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX	
DR	WPI; 1999-527416/44.
XX	
XX	New polypeptide comprising antigenic portions of M. tuberculosis
PT	
XX	Claim 4; Page 304; 323pp; English.
PS	
XX	
CC	This invention describes novel recombinant antigens and their encoding
CC	nucleic acids derived from Mycobacterium tuberculosis. The novel
CC	polypeptides are useful for detecting M. tuberculosis infection in a
CC	biological sample by detecting antibodies which bind with the
CC	polypeptides, and are useful as vaccines for immunizing against
CC	M. tuberculosis infection. The new detection methods are needed as
CC	current vaccination strategies do not provide 100% immunity.
XX	
SQ	Sequence 1166 BP; 117 A; 371 C; 121 G; 207 T; 350 other;

Db	563	tytnncamcmstctntmtmaactctkcccggycknmyctctctckccaynmaaccckkty	622
Qy	508	tccctttaaagcatctatcagggggaagggcccttccggcaactcccttccggcaac	567
Db	623	cywcnwycmyckckcagwyknmctccwaactctmyntttctctcnkcccmkaccknttct	682
Qy	568	tacttcccgctagtcattcttaccaccccatccacgcgcaaacgctcccttaccctaccgcgtt	627
Db	683	cwcscccccacakaymcyawcmntcmctckacscocyynycmmmcwcmctcwtw	742
Qy	628	gatttcccttttctccaccacaaagcgcgggggagcagcgcgggaaacgcgatc	685
Db	743	nakeancnttcttctctctmymtmackcwnntcnccksgaccyctctcactkmkckcm	800
RESULT 13			
AAx90924/c			
XX	ID	AAx90924 standard; DNA; 1925 BP.	
XX	AC	AAx90924;	
XX	DT	17-JAN-2000 (first entry)	
XX	DE	Epstein Barr Virus Nuclear Antigen 1 (EBNA 1) DNA.	
KW	KW	Epstein Barr Virus Nuclear Antigen 1 DNA; plasmid pCMVBNA; EBNA 1;	
KW	KW	episome; transfection; origin of replication; EBV orip; receptor;	
KW	KW	eucaryotic host cell; recombinant cell line; ion channel; gene therapy;	
KW	KW	multiple gene expression; transporter protein; transcription factor;	
KW	KW	adhesion molecule; antisense therapy; gene amplification;	
KW	KW	cell immortalisation; ds.	
OS	OS	Epstein-barr virus.	
XX	Key	Location/Qualifiers	
FT	CDS	1..1925	
FT	FT	/*tag= a	
FT	FT	/product= "EBNA 1"	
FT	FT	/transl_except= (pos:799..800, aa:Gly)	
FT	FT	/note= The sequence is described throughout the	
FT	FT	specification as being 1925 nucleotides long, but a	
FT	FT	sequence of only 1925 bp has been given in figure 2"	
XX	XX	W09947647-A1.	
XX	XX	23-SEP-1999.	
XX	XX	12-FEB-1999; 99WO-US03307.	
XX	XX	18-MAR-1998; 98US-0040961.	
XX	XX	06-AUG-1998; 98US-0130114.	
XX	PA	(PHAR-) PHARMACOPEIA INC.	
XX	PI	Damaaj BB, Horlick RA, Robbins AK;	
XX	XX	WPI; 1998-610610/52.	
DR	P-PSDB	AAY28843.	
XX	XX	New method for expressing genes from recombinant eukaryotic cells,	
PT	PT	useful for gene therapy	
XX	PS	Claim 24; Fig 2; 86pp; English.	
XX	XX	The present sequence is a DNA encoding Epstein Barr Virus Nuclear	
CC	CC	Antigen 1 (EBNA 1), which is obtained from commercially available	
CC	CC	plasmid pCMVBNA. EBNA 1 protein is used to stably maintain episomes	
CC	CC	containing EBV origin of replication (orip) and a gene encoding	
CC	CC	protein or RNA of interest. Eucaryotic host cells expressing EBNA 1	
CC	CC	protein are transfected with these episomes to produce recombinant	
CC	CC	cell lines expressing multiple genes of interest. This provides a	
CC	CC	rapid and reliable method of stably expressing multiple genes in	
CC	CC	transfected cells. The episomes are useful in the transfection of genes	

PI Aurias A, Delattre O, Desmaze C, Melot T, Peter M;
PI Ploucastel B, Thomas G, Zucman J;

```
XX WPI; 1993-386580/48.
DR P-PSDB; AAR44555.
XX
XX New nucleic acid of EWS gene and its hybrid(s) - contg. gene
PT sequence involved in chromosomal trans-location, also derived
PT mRNA, probes, fusion proteins etc., for diagnosis and treatment
PT of Ewing sarcoma and melanoma
XX
XX Disclosure: Fig 6; 123pp; French.
XX
XX The probes 22RR3 and 22RR12 were used to screen a human foetal
CC brain cDNA library (Stratagene cat.# 936206). The clone BFIAC5 was
CC identified and sequenced. It represents the entire coding region
CC and 3'-UTR of the Ews gene.
XX
XX Sequence 2371 BP; 639 A; 587 C; 659 G; 486 T; 0 other;
SQ
```

Query Match 2.8%; Score 37.4; DB 14; Length 2371;
Best Local Similarity 47.0%; Pred. No. 0.54;
Matches 116; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
QY 534 gaagcccttcggcaactcccttcggcaactactctcccgtaagtaacctttcacc 593
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1969 GACGGTGTCTGCTTTATCCATTTTCCAGGTCTCCACGTCCTCTTCTCTCCCA 1910
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 594 cccatccagcccaacgtccctacacctcacgcttgatttcccttcttcccaaaagc 653
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1909 TCTGTTCATCAAGGTCTCAGGGGGCCCCCAGGGCCACCTCGTCTCTCCACCAAGC 1850
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 654 agccgggggagcagccggtaaccgatactcctgacctgttgtagcggggcatcgg 713
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1849 CACCTCGGTCCATGCCCGGCCACGACGAGCCACCTCTGTCTCCACCCAGCGCCACCTC 1790
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 714 ggccggaattcgtggccttcaccagccgcgctaatacagaggggtgctgatgacac 773
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1789 TGAACATTCCACCGGACCCACCATCATGAGGCCACCTCTTCTCTCCCGCATGCCAC 1730
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 774 cattacc 780
   ||| |||
Db 1729 CAGGGCC 1723
```

Search completed: October 21, 2001, 11:33:28
Job time: 2576 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2001, 10:45:17 ; Search time 992.24 Seconds
 (without alignments)
 12803.994 Million cell updates/sec

Title: US-09-596-958-1
Perfect score: 1344
Sequence: 1 atgtcaattcttaagcttaa.....acctgaagggtgctgaatga 1344

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database :
EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estrol1:*
83: em_estrol2:*
84: em_estrol3:*
85: em_estrol4:*
86: em_estrol5:*
87: em_estrol6:*
88: em_estrol7:*
89: em_estrol8:*
90: em_estrol9:*
91: em_estrol10:*
92: em_estrol11:*
93: em_estrol12:*
94: em_estrol13:*
95: em_estrol14:*
96: em_estrol15:*
97: em_estrol16:*
98: em_estrol17:*
99: em_estrol18:*
100: em_estrol19:*
101: em_estrol20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	Count				
1	63.8	4.7	469	152	BG310823	BG310823	ro43d10.y
2	47	3.5	910	219	CNS0080N	AL065629	Drosophill
3	43.8	3.3	959	219	CNS008MZ	AL052079	Drosophill
4	43.4	3.2	844	219	CNS00852P	AL052079	Drosophill
5	42.8	3.2	782	80	BF280877	BF280877	GA_Eb004
6	42.8	3.2	782	153	BG447322	BG447322	GA_Eb004
7	41.8	3.1	1360	170	BF866129	BF866129	963066G03
8	41.6	3.1	538	105	AL514267	AL514267	AL514267
9	41.6	3.1	870	165	BE258873	BE258873	601116824
10	41.6	3.1	975	220	CNS02AKBD	AL201298	Tetraodon
11	41.2	3.1	391	152	BG310856	BG310856	ro44a03.y
12	41.2	3.1	495	17	A2101143	A2101143	qf64g10.x
13	41.2	3.1	534	112	AW173078	AW173078	xj82h08.x
14	41.2	3.1	563	19	A3336864	A3336864	qx86h08.x
15	41	3.1	467	106	AL566523	AL566523	AL566523
16	40.8	3.0	669	228	AQ447577	AQ447577	mgxb0008E
17	40.8	3.0	977	219	CNS00U0X7	AL076850	Drosophill
18	40.4	3.0	812	220	CNS02JGO	AL200193	Tetraodon
19	40.2	3.0	522	111	AW090025	AW090025	xc89d09.x
20	40.2	3.0	551	11	AA7711081	AA7711081	vt16c12.r
21	40.2	3.0	990	221	CNS03CU0	AL238257	Tetraodon
22	40.2	3.0	1101	219	CNS00EDF	AL068951	Drosophill
23	40	3.0	390	14	AF960471	AF960471	w663a05.s
24	39.8	3.0	667	168	BF700488	BF700488	602128842
25	39.8	3.0	779	227	AQ331220	AQ331220	nbxb00049E
26	39.8	3.0	860	153	BG394463	BG394463	602456990
27	39.6	2.9	539	251	AZ090203	AZ090203	RPGT-24-1
28	39.6	2.9	793	256	B18549	B18549	T5A19-Sp6.T
29	39.6	2.9	995	219	CNS018MR	AL095333	Drosophill
30	39.6	2.9	1072	221	CNS03EN5	AL240602	Tetraodon
31	39.4	2.9	878	137	BE572645	BE572645	601329421
32	39.2	2.9	968	147	BF341280	BF341280	602013214
33	39.2	2.9	1095	220	CNS02762	AL184292	Tetraodon
34	39	2.9	560	120	AW732191	AW732191	ba06f10.x
35	39	2.9	619	222	FR0007731	Z91541	F.rubripes
36	39	2.9	667	108	AL446606	AU144606	AU144606
37	39	2.9	856	220	CNS023HB	AL197412	Tetraodon
38	39	2.9	997	168	BF683247	BF683247	602139267
39	39	2.9	1101	219	CNS002B1	AL097447	Drosophill
40	38.8	2.9	398	152	BG356477	BG356477	Em1_23_A0
41	38.8	2.9	455	137	BE601363	BE601363	P11_92_A1
42	38.8	2.9	568	154	BG463697	BG463697	Em1_50_H0
43	38.8	2.9	573	119	AW679673	AW679673	WS1_30_E0
44	38.8	2.9	597	152	BG323139	BG323139	Em1_15_D0
45	38.8	2.9	626	152	BG355747	BG355747	Em1_19_A0

ALIGNMENTS

RESULT		1				
LOCUS	BG310823					
DEFINITION	Heterodera glycines J2 pAMP1 v2.1 Chiapelli McCarter ro43d10_v1 Heterodera glycines cDNA 5' similar to TR:Q9M6Z2 Q9M6Z2 PECTATE LYASE PRECURSOR ; mRNA sequence.	469 bp mRNA EST	OS-APR-2001			
ACCESSION	BG310823					
VERSION	BG310823.1	GI:13112624				
KEYWORDS	EST.					
SOURCE	Heterodera glycines.					
ORGANISM	Heterodera glycines. Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida: Tylenchina; Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.					
REFERENCE	1 (bases 1 to 469)					
AUTHORS	McCartor,J., Clifton,S., Chiapelli,B., Pope,D., Martin,J., Wylie,T., Dante.M., Marra,M., Hillier,L., Kucaba.T., Theising.B., Bowers.Y., Gibbons,M., Ritter.E., Bennett.J., Franklin,C., Tsagareishvili.R.					


```

Query Match          3.1%; Score 41.6; DB 105; Length 538;
Best Local Similarity 43.1%; Pred. No. 0.61;
Matches 91; Conservative 17; Mismatches 103; Indels 0; Gaps 0;

QY 539 ccccttcgggaactcccttcggaactactctccgctcagtcactcttccaccccat 598
      :||| || | |||| || | |||| || | |||| || | |||| || | ||||
Db 100 MCCACMCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 159
      :||| || | |||| || | |||| || | |||| || | |||| || | ||||
QY 599 ccacgcaagctccctactacacgctgattctctctctctctccacaaagagcg 658
      || | || | |||| || | |||| || | |||| || | |||| || | ||||
Db 160 CCCCCCCCCCCCCCCCCCCCCCTTCCTCTTTCCTCCCCCCCCGCCGCCGCCGCC 219
      || | || | |||| || | |||| || | |||| || | |||| || | ||||
QY 659 gggcgagcagcggtaacgatcatcctgacccctgttgtagcggggcacgcggcg 718
      || | || | |||| || | |||| || | |||| || | |||| || | ||||
Db 220 GGGCCCCCCCCCANNCCNGGGCCCCCCCCCNCNCCCCYBCCCCCGGGGGGGG 279
      || | || | |||| || | |||| || | |||| || | |||| || | ||||
QY 719 gaaatcggtggtccctcaccagcgcgggcg 749
      || | || | |||| || | |||| || | |||| || | |||| || | ||||
Db 280 GSCCGGGGGGSCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 310
      || | || | |||| || | |||| || | |||| || | |||| || | ||||

RESULT 9
BE258873
LOCUS      BE258873      870 bp      mRNA      EST      13-JUL-2000
DEFINITION 601116824F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3357533 5',
            mRNA sequence.
ACCESSION  BE258873
VERSION    BE258873.1 GI:9129368
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 870)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgaabsr@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: L1CM164 row: o column: 06.
FEATURES   Location/Qualifiers
            source
              1..870
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3357533"
                /clone_lib="NIH_MGC_16"
                /tissue_type="retinoblastoma"
                /lab_host="PH10B (phage-resistant)"
                /note="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GGACGAG(G). Library constructed by Ling Hong
                in the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."
            BASE COUNT      43 a 716 c 67 g 44 t
            ORIGIN

Query Match          3.1%; Score 41.6; DB 165; Length 870;
Best Local Similarity 50.5%; Pred. No. 0.7;
Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 449 ttggccaactgtgtaggggcaacacagtcctcttcggtactcttcttcacgtggtt 508
      | | | | | ||| | | | | | | | | | | | | | | | | | | | | |

```

```

Db 256 TCGCCGCCCTCGCTCGCGGCCGCCGCCGCCACATCGCGCGGCCGCCGCTTCTCGCGCGCCGCT 315
QY 509 ccccttttaacgatctatcagggggaagccctctccggaactcccttcgggaact 568
      | | | | | ||| | | | | | | | | | | | | | | | | | | | | |
Db 316 CGCTCCGCCCTCGCTTCACCCCCCTCGGCCCTCGCGGCCGCCGCCGCCGCCGCCGCC 375
      | | | | | ||| | | | | | | | | | | | | | | | | | | | | |
QY 569 actctccgctcagtaactcttctcaccgccatccacgcaacgctccctacacccgttg 628
      | | | | | ||| | | | | | | | | | | | | | | | | | | | | |
Db 376 CFTCCGCCCTCGGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 435
      | | | | | ||| | | | | | | | | | | | | | | | | | | | | |
QY 629 atttcctcttcttcacac 648
      | | | | | ||| | | | | | | | | | | | | | | | | | | | | |
Db 436 CCGCGCTCGGCCGCCGCCGCC 455
      | | | | | ||| | | | | | | | | | | | | | | | | | | | | |

RESULT 10
CNS02KBD/C
LOCUS      CNS02KBD/C      975 bp      DNA      GSS      14-MAY-2000
DEFINITION Tetraodon nigroviridis' genome survey sequence PUC-Ori end of clone
            145K19 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL201298
VERSION    AL201298.1 GI:7859643
KEYWORDS   GSS; genome survey sequence.
SOURCE     Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE  1 (bases 1 to 975)
AUTHORS   Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
            Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
            Weissenbach, J.
TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 975)
AUTHORS   Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
            Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
            Saurin, W. and Weissenbach, J.
TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
JOURNAL   Unpublished
REFERENCE  3 (bases 1 to 975)
AUTHORS   Genoscope.
TITLE     Direct Submission
JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT   This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES   Location/Qualifiers
            source
              1..975
                /organism="Tetraodon nigroviridis"
                /db_xref="taxon:99883"
                /clone="145K19"
                /clone_lib="G"
                /note="Genoscope sequence ID : COAG145AF10SP1-end ;
                PUC-Ori"
            BASE COUNT      182 a 128 c 348 g 160 t 157 others
            ORIGIN

Query Match          3.1%; Score 41.6; DB 220; Length 975;
Best Local Similarity 32.6%; Pred. No. 0.72;
Matches 71; Conservative 49; Mismatches 98; Indels 0; Gaps 0;

QY 539 ccccttcgggaactcccttcggaactactctccgctcagtcactcttccaccccat 598
      :||| || | |||| || | |||| || | |||| || | |||| || | ||||
Db 594 SCCCCMCSGCMTCYSCCCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 535
      :||| || | |||| || | |||| || | |||| || | |||| || | ||||
QY 599 ccacgcaagctccctactacacgctgattctctctctctccacaaagagcgcg 658
      || | || | |||| || | |||| || | |||| || | |||| || | ||||

```



```

Db 312 CCTCTTCAGCTCCTCTCTCCAGCTCCTCTCTCCATATTCAGCCTCTT 371
Qy 599 ccacgcaagtcctccctacaccggttgatttccttcttccacc 648
Db 372 CGACCTCTTCTCTCTCTCCAGCTCCTCTCTCCATATTCAGCCTCTT 421

RESULT 13
AW173078
LOCUS
DEFINITION
XJ82H09.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2663777 3' similar to contains element MSRI repetitive
element ;, mRNA sequence.
ACCESSION
AW173078
VERSION
AW173078.1 GI:6439026
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 534)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 466.
FEATURES
source
1..534
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2663777"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI-CGAP_GC81) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 102 a 174 c 68 g 190 t
ORIGIN

Query Match 3.1%; Score 41.2; DB 112; Length 534;
Best Local Similarity 60.9%; Pred. No. 0.79;
Matches 67; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 539 ccccttcggcaactcccttcggcaactactctcccgtagtaccttctccaccat 598
Db 312 CCTCTTCAGCTCCTCTCTCCAGCTCCTCTCTCCATATTCAGCCTCTT 371
Qy 599 ccacgcaagtcctccctacaccggttgatttccttcttccacc 648
Db 372 CGACCTCTTCTCTCTCTCCAGCTCCTCTCTCCATATTCAGCCTCTT 421

RESULT 14
AI336864
LOCUS
DEFINITION
qx86h08.x1 NCI-CGAP_GC6 Homo sapiens cDNA clone IMAGE:2009439 3'
similar to contains element MSRI repetitive element ;, mRNA

```

```

sequence.
ACCESSION
AI336864
VERSION
AI336864.1 GI:4073791
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 563)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1179 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 448.
FEATURES
source
1..563
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2009439"
/clone_lib="NCI-CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Plasmid DNA from the normalized library
NCI-CGAP_GC4 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (clones 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 113 a 176 c 67 g 207 t
ORIGIN

Query Match 3.1%; Score 41.2; DB 19; Length 563;
Best Local Similarity 60.9%; Pred. No. 0.8;
Matches 67; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 539 ccccttcggcaactcccttcggcaactactctcccgtagtaccttctccaccat 598
Db 354 CCTCTTCAGCTCCTCTCTCCAGCTCCTCTCTCCATATTCAGCCTCTT 413
Qy 599 ccacgcaagtcctccctacaccggttgatttccttcttccacc 648
Db 414 CGACCTCTTCTCTCTCTCCAGCTCCTCTCTCCATATTCAGCCTCTT 463

RESULT 15
AL566523/c
LOCUS
DEFINITION
AL566523 LTI_FL013_FBrnl Homo sapiens cDNA clone CS0DF020YA13 3
prime, mRNA sequence.
ACCESSION
AL566523
VERSION
AL566523.1 GI:12918965
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

Search completed: October 21, 2001, 11:30:56
Job time: 2739 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2001, 10:17:57 ; Search time 1309.91 seconds
(without alignments)
15870.293 Million cell updates/sec

Title: US-09-596-958-1
Perfect score: 1344
Sequence: 1 atgtcaattcttaacgcttaa.....acctgaagggtgctgaatga 1344

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb_bal:*
 - 2: gb_ba2:*
 - 3: gb_ba3:*
 - 4: gb_in1:*
 - 5: gb_in2:*
 - 6: gb_in3:*
 - 7: gb_om:*
 - 8: gb_ov:*
 - 9: gb_pat1:*
 - 10: gb_pat2:*
 - 11: gb_phi:*
 - 12: gb_pli:*
 - 13: gb_pl2:*
 - 14: gb_pl3:*
 - 15: gb_pl4:*
 - 16: em_bal:*
 - 17: em_ba2:*
 - 18: em_fun:*
 - 19: em_htgo_hum:*
 - 20: em_htgo_inv:*
 - 21: em_htgo_rod:*
 - 22: em_htg_hum1:*
 - 23: em_htg_hum2:*
 - 24: em_htg_hum3:*
 - 25: em_htg_hum4:*
 - 26: em_htg_hum5:*
 - 27: em_htg_hum6:*
 - 28: em_htg_hum7:*
 - 29: em_htg_hum8:*
 - 30: em_htg_inv1:*
 - 31: em_htg_inv2:*
 - 32: em_htg_other:*
 - 33: em_htg_rod:*
 - 34: em_hum1:*
 - 35: em_hum2:*
 - 36: em_hum3:*
 - 37: em_hum4:*
 - 38: em_hum5:*
 - 39: em_hum6:*
 - 40: em_hum7:*
 - 41: em_in:*
 - 42: em_om:*
 - 43: em_or:*

- 44: em_ov:*
- 45: em_pat:*
- 46: em_ph:*
- 47: em_pl:*
- 48: em_ro:*
- 49: em_sts:*
- 50: em_sy:*
- 51: em_un:*
- 52: em_vi:*
- 53: gb_sts1:*
- 54: gb_sts2:*
- 55: gb_sts3:*
- 56: gb_sy:*
- 57: gb_un:*
- 58: gb_vil:*
- 59: gb_v12:*
- 60: gb_htg1:*
- 61: gb_htg2:*
- 62: gb_htg3:*
- 63: gb_htg4:*
- 64: gb_htg5:*
- 65: gb_htg6:*
- 66: gb_htg7:*
- 67: gb_htg8:*
- 68: gb_htg9:*
- 69: gb_htg10:*
- 70: gb_htg11:*
- 71: gb_htg12:*
- 72: gb_htg13:*
- 73: gb_htg14:*
- 74: gb_htg15:*
- 75: gb_htg16:*
- 76: gb_htg17:*
- 77: gb_htg18:*
- 78: gb_htg19:*
- 79: gb_htg20:*
- 80: gb_htg21:*
- 81: gb_htg22:*
- 82: gb_htg23:*
- 83: gb_htg24:*
- 84: gb_htg25:*
- 85: gb_pr1:*
- 86: gb_pr2:*
- 87: gb_pr3:*
- 88: gb_pr4:*
- 89: gb_pr5:*
- 90: gb_pr6:*
- 91: gb_pr7:*
- 92: gb_pr8:*
- 93: gb_pr9:*
- 94: gb_rol:*
- 95: gb_ro2:*
- 96: gb_in4:*
- 97: gb_pr10:*
- 98: em_ba3:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1344	100.0	2708	3	U94513
c	2	1344	100.0	9985	2 EADSPAB
c	3	1329.6	98.9	4291	1 AF083620
c	4	414	30.8	8143	2 EAU97504
5	177	13.2	1833	1	AF037983
6	155.2	11.5	1729	1	AF005221
7	155.2	11.5	20052	16	AF232006
8	155.2	11.5	52498	2	AF232004

9	95.6	7.1	1245	1	AB011839
c 10	91	6.8	47739	1	AF017113 Bacillus
11	91	6.8	209510	2	BSU00018
12	63.2	4.7	1214	2	BJ23237980
13	62.2	4.6	884	14	FSOPELAB
14	61.4	4.6	40549	3	SC2G38
15	50.2	3.7	1419	14	FSU13050
16	45.8	3.4	7218	10	I66494
17	43.4	3.2	1332	14	FSU13049
c 18	43.2	3.2	7218	10	I66494
19	42.8	3.2	95432	14	ATT9E8
20	42.8	3.2	199634	13	ATCHRIV36
21	42.4	3.2	2185	14	FSOPELAA
22	42.4	3.2	213685	83	AP002760
c 23	42.2	3.1	209119	76	AC083818
c 24	42	3.1	237678	94	AC055818
c 25	41.8	3.1	152528	64	AC016063
c 26	41.8	3.1	183099	69	AC025288
c 27	41.6	3.1	5664	1	AE005748
c 28	41.6	3.1	68326	72	AC055793
c 29	41.2	3.1	2805	7	AF286474
c 30	41.2	3.1	82024	67	AC023210
c 31	41.2	3.1	194834	75	AC078873
c 32	40.8	3.0	203587	60	AC007491
c 33	40.6	3.0	221840	63	AC012618
c 34	40.2	3.0	169434	90	AL359205
c 35	40.2	3.0	171588	67	AC023198
c 36	40.2	3.0	198692	78	AC090530
c 37	40	3.0	135039	86	AC006060
-38	39.8	3.0	156975	81	AL442123
c 39	39.8	3.0	170396	81	AL512424
c 40	39.8	3.0	192730	97	HU95743
c 41	39.4	2.9	88036	97	HUAC003119
c 42	39.2	2.9	162342	80	AL359966
-43	39	2.9	166735	79	AL355861
c 44	38.8	2.9	861	14	CR6GCR4
c 45	38.8	2.9	90832	85	AC003065

ALIGNMENTS

RESULT	1
U94513	
LOCUS	U94513 2708 bp DNA BCT 03-OCT-1998
DEFINITION	Erwinia amylovora disease-specific operon, partial sequence; and hrpW operon, complete sequence.
ACCESSION	U94513
VERSION	U94513.1 GI:3414585
KEYWORDS	type III protein secretion; Hrp pathway; hypersensitive response; pectate lyase.
SOURCE	Erwinia amylovora.
ORGANISM	Erwinia amylovora.
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Erwinia.
AUTHORS	Kim, J.F., Zumoff, C.H. and Beer, S.V.
TITLE	hrpW, a new harpin of Erwinia amylovora, is a member of a family of pectate lyases
JOURNAL	Phytopathology 87, 552 (1997)
REFERENCE	2 (bases 1 to 2708)
AUTHORS	Kim, J.F. and Beer, S.V.
TITLE	HrpW of Erwinia amylovora, a new harpin that contains a domain homologous to pectate lyases of a distinct class
JOURNAL	J. Bacteriol. 180 (19), 5203-5210 (1998)
MEDLINE	98422475
REFERENCE	3 (bases 1 to 2708)
AUTHORS	Kim, J.F. and Beer, S.V.
TITLE	Direct Submission
JOURNAL	Submitted (18-MAR-1997) Plant Pathology, Cornell University, 334 Plant Science Bldg., Ithaca, NY 14853, USA
FEATURES	Location/Qualifiers
source	1..2708

/organism="Erwinia amylovora"
 /strain="Ba321; ATCC 49947"
 /db_xref="taxon:552"
 /db_xref="ATCC:49947"
 /note="a strain that infects apples, pears, and other
 pomaceous plants"
 complement(<1..784)
 /gene="disease-specific operon"
 complement(<1..729)
 /gene="dspe"
 complement(<1..714)
 /gene="dspe"
 /note="Hrp-secreted pathogenicity/avirulence protein
 similar to Pseudomonas syringae AvrE"
 /codon_start=1
 /transl_table=11
 /product="dspe"
 /protein_id="AAC62315.1"
 /db_xref="GI:3414587"
 /translation="MELKSLGTEHKAHVTAHNPVGHVALQOQSSSSPQNAASL
 AAEKNGKMPRIHOPSTADGISAHQKKSFLRGLGTFKFSRAPOGPGTTHS
 KGATRLDLDADGCTOHEAAPDAARLTSGGVKRRMDDMAGRPVKGSGGDKVP
 TQKRHLNFGQMRQWLSKMAHPASANAGDRLQHSPPHPCSHHEIKKEPVGSTK
 ATTAHADRVETAEQEDDDSEF"
 complement(724..729)
 /gene="dspe"
 complement(755..784)
 /gene="disease-specific operon"
 /note="HrpL-dependent promoter consensus sequence; hrp
 box"
 825..854
 /gene="hrpW operon"
 /note="HrpL-dependent promoter consensus sequence; hrp
 box"
 825..2692
 /gene="hrpW operon"
 900..905
 /gene="hrpW"
 900..2255
 /gene="hrpW"
 /gene="harpin gene"
 912..2255
 /gene="hrpW"
 /function="elicit the hypersensitive reaction"
 /note="Hrp-secreted; C-terminal domain similar to class
 III pectate lyases; high glycine and serine content; lacks
 cysteine"
 /codon_start=1
 /transl_table=11
 /product="HrpW"
 /protein_id="AAC62314.1"
 /db_xref="GI:3414586"
 /translation="MSLTLTNNNTSSPGLFQSGDNGLGHNNANSALGQPIDRQTI
 EQAQLLAELLKSLSPQNAATGAGNDQTTGVNAGNLRKGTAGTTPQSDSN
 MLSMNGNGLDQAITPDGQGGQIDGNPLKMLKLIARMMDGSDGFGPGTGNNNA
 SSGTSSGGSPFNLDGSGKAPSGNSPGNSPVSTFSPSTPSPLDPPSPPTKA
 AGSTPYTHDPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGOVFDGKGTFTA
 GSELGDGQSGENKQPLEIDGASLKNVTMGDDGADGCIHLGDKAKINLHVTNVEDA
 ITVPNSAGKSHVEITNSSEFHASDKILQNLADTNLSVDNVRKAKDFTGRTNGGOO
 GNMDNLNLSHISAEDGKFSFKVKSDEGLNVNTSDISLGDVENHYKVPMSANLVAE"
 2242..2247
 2252..2584
 /note="putative; OrfC; high glutamine content"
 /codon_start=1
 /transl_table=11
 /product="potential HrpW-specific chaperone"
 /protein_id="AAC62316.1"
 /db_xref="GI:3414588"
 /translation="MMKSAEIPAESWVDVNSQOHLSPAQYQAFQKAIQVOORLH
 QVLSRPMQPGKFFELDSFVDSLHADFLIGSDSDSNVGTAWWITRYLADRLLELQ
 QLOQQRGL"
 2655..2692
 /gene="hrpW operon"

terminator

RESULT 2

LOCUS

ACCESSION

KEYWORDS

SOURCE
ORGANISM

AUTHORS

✓

JOURNAL

REFERENCE

TITLE

1

REFERENCE

TITLE

.....

AUTHORS

JOURNAL

COMMENT

source

```
/clone="pMAB37"
/lab_host="pear and apple trees"
/complement(1605..3035)
/gene="hrpW"
/complement(1605..2948)
/gene="hrpW"
/codon_start=1
/transl_table=11
/product="HrpW protein"
/protein_id="CAA74158.1"
/db_xref="GI:3150150"
/db_xref="SPTREMBL:O54508"
/translation="MSILTLNNNTSSPLFQSGDNGLGHNNANSLAQPIQDRITTI
EQMAQLLESLKSLSPGAGNAGTGGNAGTGGNAGLNGRGTAGTTPQSDSON
MLSENGNLQDAITPDQGGQIGDNPDLKAMLLIARMMDGSDQFQPGCTGNNSA
SSGTSSSGSPNDLSGGKAPSGNSPNSVPSFSPSTPTSPDLPSPTKA
AGGSTPVYDHPDPVGSAGIAGNSVAFTSAGANQTVLHDTITVRAGQVFDKGTFTA
GSELGDGSGENKPLFLIEDGASLKNVTMGDGDAGIHLGYDADKIDMLNVTGEDA
ITVPNSAGKXSHVEITRSEFHASDKTLQNLADTNLSVDNFKADFGTFVYRTNGGQ
GNWDLNLSHISAEDGKFFSVKSDSEGLNVNTSDISLGDVENHYKVPMSANLKVAE"
/complement(3008..3035)
/gene="hrpW"
/promoter
3076..3101
/gene="dsbA"
3076..8662
/gene="dsbA"
3146..8662
/gene="dsbA"
/function="involved in pathogenicity"
/codon_start=1
/transl_table=11
/product="DspA protein"
/protein_id="CAA74156.1"
/db_xref="GI:2695858"
/db_xref="SPTREMBL:O54581"
/translation="MELKSLGTEHKAHVHTAAHNPVGHGVALQCGSSSSSSPQNAASL
AAEGRNKMRIPIHPSAADGISAHQOKFSFLRGCLGTFKFSRSAPQOPGTTHS
KGATURLDNRDQTEHEAAPAARLTRSGGVKRRNMDDMPVMVKGSGEDKVP
TQQRHLQNNFQMRQDTMLSKMHPASANAGDRLOHSPHPIPGSHHEIKESPVGSTK
ATTAHADRETAQEDDSEFOQLHOORLARENPPOPKLGVATPI SAROPKLTA
AESVLEGTDTTQSPKPSMLKSGAGVTPLAITLDKGLQLOLPNDPNTLLKQTL
CKDQHYLAHASSDGSQHLLDNKGHLFDIKSTATSYSLVNSHPHBEIKGLAQGT
GKSVSDHGLSGVSGTQSHNKTMLSQGEAHRSLGLIWOHPAGAARPOGESIRLH
DSKTHILPELQSGADKTHSLSQADGKLKALKDNRTLQNLSDNKSSEKLVDKI
KSYVDQRGVAILTDTPEGRHMSIMPSLDASPEHSISLSLHFADAHQGLHDSLE
AQSAIVSHGRVAVDSEKLFSAALPKQDGNELKMKAMPOHALDEHGHDOISGFF
DDHGQALNALVKNPQROHACPLGNHQHFGPWNLTDAVTDNOLGLHHTYEPHEIL
DMGSLKGLAQEGKLHYFDQLTKWGTAESECKQLKGLDGAAYLLKDKGPKRLNINQ
STSIKSGTENRFLPHVRNRPEDGALQGLNKKDAQAMAVIGYNNKYLALTEKGDIR
SFQIKPGTQQLERPAQTLUREGISGELKDIHVDHKNDLYALTHEGEVHPQREAWQNG
AESSWHKRALPQSESKLSLSDMSHEHKPIATFEDGSQHLKAGWHAYAAPERGPLA
VGTSGSQTVPNRLMQGVKGVIPGSLTVKLSAOTGMGTGAERKVSFKFSERTRAYA
FNPTMSTPRPKNAAYATQHQWQREGKLPSEMOGALIKOLDAHNVHRNAPQDLOS
KLETLDLCEHGAELLNDMKRPRDELEQASRTSVTVLQGHQVLSNGEINSEFPSPG
KALVQSPNVNSGQDLSLQQAQVHATPPSAESKLQSLMGLHFVSAGVDMSHQKEIPL
GRQDPNDKTLASRLILDVTYIGELHEADKALVSDHDPDADQIKOLROQFDTLR
EKRESNPVKTYTDMGFTHNKALEANYDAVKAFNAFKKEHGVNLTTRTVLESQGS
ELAKKLTLLSLSDGSEMSFSRSYGGVGVTFVFTLSKKVPPVPIPGAGITLDRAVN
LSFSTSGGLANVSGRDGSGVSNIMVATGHQVMPYMTGKTSAGNASDMLSAKHKISP
DLRIGASVGTLOGLQNSLKEFLTEDELPGFIHGLTHGTLTAPLLELQKHQHKQK
SLRTFSVDTSLNDLURGAILNEDGSKPNGVARTVASGLSASNLAAAGRSRSTTSGQ
FGSTTSANNRPRTFLNGVAGANTALTAALGVAHSSTHEKPVGIFPAFTSTNVSAAL
DNRTSQSLSBELKAPYTSNDISELTGLTHGFKDSATKMLAALKELDNAKAEQL
HILQOHSKADVGVGDERYEVRNLKLVIRQOADSWSMELGSSASHSTYNNLRINN
DGIIVELLHKHFDAAALPASSAKRLGEMMNDPALKDIKQLOSTPFSASVSMLKDGIL
REOTEKATILDGKGVREEVGLFQDRNNLRVKSVSQSVSKSEGFNTALLGTSNSA
AMSHERNIGTINFTYQDQNTPRRFTLEGGIAQANPQVASALTDLKREGLEMK"
8675..8701
/gene="dsbB"
8675..9143
/gene="dsbB"
8724..9143
/gene="dsbB"
/promoter
8675..8701
/gene="dsbB"
8675..9143
/gene="dsbB"
8724..9143
/gene="dsbB"
```

```
/function="involved in pathogenicity"
/codon_start=1
/transl_table=11
/product="DspB protein"
/protein_id="CAA74157.1"
/db_xref="GI:2695859"
/db_xref="SPTREMBL:O54621"
/translation="MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAALVE
VPOHSDLLHCRITIEADPQTSITLYSMLQLNFMMAARMCWLALDELHNVRLCFQ
SLEHLDEASFSDIVSGFIEHAAREVYIAQLEDESSAA"
BASE COUNT 2570 a 2562 c 2726 g 2127 t
ORIGIN
Query Match 100.0%; Score 1344; DB 2; Length 9985;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgtcaattcttacgcttaacaacaataacctgctctcgcgggtctgttccagtcggg 60
Db 2948 ATGTCAATTTCTTACGCTTAACAACAATACCTCGTCTCGCGGGTCTGTTCAGTCCGGG 2889
QY 61 ggggacaacgggcttgggtggtcataatgcaaatcttgggtggggaacacccatcgat 120
Db 2888 GGGGACAACGGGCTTGGTGTCTAATATGCAAAATTCGCGTTGGGGCAACACCCATCGAT 2829
QY 121 cggcaaacattgagcaaatggctcaattattggcggaactgttaagtcactgtatcg 180
Db 2828 CGGCAAACCATTTGAGCAAAATGGCTCAATTTATTGGCGGAACCTGTTAAGTCACGTCTATCG 2769
QY 181 ccacaatcaggtaatgcggcaacccggagccggtggcaatgaccagactacagagtggt 240
Db 2768 CCACAATCAGGTAATGCGGCAACCGGAGCGGTGCAATGACCAGACTACAGGAGTGGT 2709
QY 241 aacgctggcgccctgaacgagcaaaaagcagcaggaacacactccgcagtcctacag 300
Db 2708 AACGTGGCGGCTTGAACGGAGCAAAAGSCACAGCAGGAAACCACTCCGCGAGTCTCACAGT 2649
QY 301 cagaacatgctgagtgagatggcgcaacggcggtggatcagggccatcacgcccagatg 360
Db 2648 CAGAATGCTCAGTGAGATGGGCAACACAGCGGTGGATCAGGCCATCAGGCCCATGCG 2589
QY 361 cagggcgcgggcagatggcggaataatccttctactgaaagccatgctgaagcttattg 420
Db 2588 CAGGCGCGGCGGAGATCGCGGATAATCTTTACTGAAAGCCATGCTGAAGCTTATTGCA 2529
QY 421 cgcatagtgacggcccaaacgacatcagtttggcgaacctggtacgggaacacacagtc 480
Db 2528 CGCATGTGGACGGCCAAAGCCATCAGTTTGGCCAACTTGGTACGGGCAACACAGTGCC 2469
QY 481 tcttcgggtacttcttcaatctggcggttccctcttcaacgactctatcagggggggaagcc 540
Db 2468 TCTTCCGCTACTTCTTCATCTGCGGTTCCCTTTTAAAGGATCTATCAGGGGGGAAGGCC 2409
QY 541 ccttcgggaacaccccttcgggaactactctccgctcgcgctcgcgctcgcgctcgcg 600
Db 2408 CCTTCGGCAACTCCCTCTCCGGAACACTACTCTCCGTCAGTACCTTCTTCACCCCATCC 2349
QY 601 acgcaacgctccctacactcaccgcttgatttcccttcttcccccacacacagccggg 660
Db 2348 AGCCCAAGCTCCCTTACCTTACCGGCTTGATTTCCCTTCTTCTCCCAACCAACAGCCGGG 2289
QY 661 ggcagcagccgggtaacgatactcctgacccctgttgtagcgcgggatcgatcggggcccga 720
Db 2288 GGCAGCAGCCGGTAACCGATCATCTGACCTCTGTTGTTAGCGCGGCATCGGGCCGGA 2229
QY 721 aattcgggtgacctcaacgacgcccgcgctaatacagaagtggtgctgcatgacaccattacc 780
Db 2228 AATTGCGTGGCTTACCAGCGCGCGGCTAATAAGACGGTGTCTGCATGACACCATTTACC 2169
QY 781 gtgaagcgggtcaggtgttgtatggcaagacaaaaccttaccgcccgttccagaatta 840
Db 2168 GTGAAGCGGGTCAAGTGTGTGTTGCAAGGACAAACCTTCCACCGCGGTTCAAGATT 2109
```

```
QY 841 ggcatggcggccagctctgaaacacagaaacccgctgtttatatactggaagacggtgcacc 900
|||||
Db 2108 GGCATGGCGGCCAGTCTGAAACACAGAAACCGCTGTTATATCTGGAAGACGCTGCCAGC 2049
QY 901 ctgaaaaacgtccaccatggcgacagcagcggggcggaatggtattcatctttacggtgatgcc 960
|||||
Db 2048 CTGAAAAACGCTACCATGGCGACGACGCGGCGGATGATTCATCTTTACGGTGATGCC 1989
QY 961 aaaaatagacaactctgacatccacacgctgggtgagagcgcgattaccgttaagccaaac 1020
|||||
Db 1988 AAATAAGACAATCTGCACGCTCAACACGCTGGGTGAGGACGCGATTACCGTTAAGCCAAAC 1929
QY 1021 agcgggggcaaaaaatccacagcttgaaatcaactaacagttctcttcagcacgcctctgac 1080
|||||
Db 1928 AGCGCGGCAAAAAATCCACGCTTGAATCACTAACAGTTCTCTTCGAGCACGCTCTGAC 1869
QY 1081 aagatcctgcagctgaatcccgatatacctacacgtgagcgttgacaacatgaaagccaaagac 1140
|||||
Db 1868 AAGATCTCTGACGTGAATCCGATACCTAACCTGAGCGTTGACAACGTGAAGGCCAAAGAC 1809
QY 1141 ttctgactctttgacgactaacgcgcgctcaacaggttaactggatctgaatctgagc 1200
|||||
Db 1808 TTTGGTACTTTTGTACGCACTAACGCGGCTCAACAGGTAACCTGGGATCTGAATCTGAGC 1749
QY 1201 catatcagcgcagaagacggttaagtctctctgttcttaaaagcgatagcaggggcgtaaac 1260
|||||
Db 1748 CATATCAGCGCACAGACGCTGAAGTTCTGCTGCTTAAAGCGATAGCGAGGCGCTAAAC 1689
QY 1261 gtcaatacagctgatctcactcgggtgtagttgtgaaacacactacaaagtgcgagtgc 1320
|||||
Db 1688 GTCAATACCACTGATATCTCACTGGGTGATGTTGAAACCACTACAAAGTGCCGATGTC 1629
QY 1321 qccaaactaaagtgcctaataa 1344
|||||
Db 1628 GCCAACTGAAGTGGCTGAATGA 1605

RESULT 3
AF083620/c
LOCUS
DEFINITION
Erwinia amylovora harpin HrpN (hrpN) gene, partial cds; potential
ORFB-specific chaperone, virulence/avirulence effector protein
homolog, probable HrpW-specific chaperone, and harpin HrpW genes,
complete cds; and Hrp-secreted pathogenicity/avirulence protein
DspE (dspE) gene, partial cds.
ACCESSION AF083620
VERSION AF083620.1 GI:7542322
SOURCE
Erwinia amylovora.
ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Erwinia.
REFERENCE
1 (bases 1 to 4291)
AUTHORS Kim,J.F., Laby,R.J. and Beer,S.V.
TITLE Comparison of the hrpN-flanking regions of two Erwinia amylovora
strains with different host specificity
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 4291)
AUTHORS Kim,J.F., Zumoff,C.H. and Beer,S.V.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1998) Plant Pathology, Cornell University, 334
Plant Science Bldg., Ithaca, NY 14853, USA
FEATURES
source
1. .4291
/organism="Erwinia amylovora"
/strain="Ea246"
/db_xref="taxon:552"
/note="pathogenic to Rubus plants"
<1..429
/gene="hrpN"
gene
<1..429
complement(2587..2592)
CDS
complement(3929..3934)
/gene="hrpW"
```

```
/note="HrpN family harpin"
/codon_start=1
/transl_table=11
/product="harpin HrpN"
/protein_id="AAF63398.1"
/db_xref="GI:7542323"
/translation="GTGICMKAGIQALINDIGTHSDSSTRSFVNKGDRAMAKEIGOFMD
QYVEFGKPYOKGPGQEVKTDKSWAKALSXPDDGHTPASMEQFNKAKGIKSAMA
GDTGNGNLQARGAGSSSLGIDAMMAGDAINNNAKLGKAA"
472..477
RBS
481..491
CDS
/note="ORFA; member of Sych family of chaperones and
chaperone-like proteins; leucine-rich"
/codon_start=1
/transl_table=11
/product="potential ORFB-specific chaperone"
/protein_id="AAF63399.1"
/db_xref="GI:7542324"
/translation="NMIIQDLGALARLEAGTSLSEAGQLCCQLVNGLEMTLEWLEQ
QNMFLVLSVGTLSGTSTLLSILAANLPHYGSDGAAGFLDKELNELLFORFQL
SSVDEAFVSACVQMEIAVKRWQKFLHGRALPTAAPRLAQPLGLTLNLAGIK"
936..941
RBS
948..2129
CDS
/note="ORFB; similar to ORFB of Erwinia amylovora Ea321,
AvrRxv of Xanthomonas campestris pv. vesicatoria, AvrA of
Salmonella enterica, and YopJ of Yersinia
pseudotuberculosis"
/codon_start=1
/transl_table=11
/product="virulence/avirulence effector protein homolog"
/protein_id="AAF63400.1"
/db_xref="GI:7542325"
/translation="MNIISLGRGYKSOAQADNASSSSSTOGSPAPTRRLRQDALPA
NYRYHSONPATPERVAARYASOASSAGPSILLRSQSHRENPRLVRFHETWQOS
PKMSRGDPLPERPEIVPKRQEKIDSYNLPRNKLDKNLYEKNKATELAKGSSSV
ALMRDKKVLPLLAENARNPGLNLHVYKRGEECYQAIKEQHKIVQOSGQPKTRAL
YPTFTGMPDHHIALDIHLRPHRPSIVGFSALGHMVDPIROGIAOGLRGAKVHMVG
RIONSEWDCIMYSLNNAKSFKHDEYTLARLHKGEKIPVPAEFFHAOSKSMVEGLPH
QDAIVTKDKGGLHAETLLHRLNLAIRADRFHACNTSIEGFRMQEIQRAEPLSAQNRK
S"
2145..2184
terminator
complement(2253..2582)
CDS
/note="ORFC"
/codon_start=1
/transl_table=11
/product="probable HrpW-specific chaperone"
/protein_id="AAF63401.1"
/db_xref="GI:7542326"
/translation="MMKSAEIPPAESWVDVYVNSQORLSPEQYQAFQRAISQVKORLQ
QVLGRPMPORGFEFLDSFVLSHADFLNVDGKAANDVDVGTAWITRYLADRLLELQ
QLEQQRG"
gene
complement(2579..3934)
CDS
complement(2579..3922)
/note="hrpW"
/gene="hrpW"
/note="HrpW family harpin; C-terminal domain similar to
class III pectate lyases"
/codon_start=1
/transl_table=11
/product="harpin HrpW"
/protein_id="AAF63402.1"
/db_xref="GI:7542327"
/translation="MSILTLNSNTSSPGLFSGGNGLGGHNANSALGQOPIDROTI
EQMAQLLAELLSLSPQSGNAATGAGNDQTTGVNAGGLNKRGTAGTTQTQSDSON
MISEMNGLDQAITPDGQGGQIDGNPLUKAWLKLIARMQDQSDQFGQPGTGNSA
SSGTSSSGSGSPNDLSGGKAGSPNSYSPVSTFSPSTPSTPSTPSTPSTPSTKA
AGGTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAISOVDKQGTFTA
GSELGDGQSENOKPLFILEDGLASLKNVTMGDDGADGILHYDAGIDNLHVTNGVEDA
ITVKPNSAGKKSHVEITNSSFEHSDKILQLNADTNLSVDNVKAKDFGTFVETNGQQ
GNWDLNLSHISEDGKSFKFSVSEGLKAVNTSDISLSDVENHYKVPMSANLKVAE"
complement(2587..2592)
RBS
complement(3929..3934)
RBS
/gene="hrpW"
```

```
promoter 3980..4005
/note="HrpL-dependent promoter consensus; hrp box"
promoter 4050..4080
/note="HrpL-dependent promoter consensus; hrp box"
gene 4105..>4291
/gene="dspe"
RBS 4105..4110
/gene="dspe"
CDS 4120..>4291
/gene="dspe"
/note="similar to AvrE of Pseudomonas syringae"
/codon_start=1
/transl_table=11
/product="Hrp-secreted pathogenicity/avirulence protein
DspB"
/protein_id="AAF63403.1"
/db_xref="GI:7542328"
/translation="MEKLSLGTETHEKAHVHTAAHNPVGHVALQGGSSSSPQNAASL
AAEKNRGEEMPRI"
BASE COUNT 996 a 1109 c 1215 g 971 t
ORIGIN

Query Match 98.9%; Score 1329.6; DB 1; Length 4291;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1335; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 atgtcaattctacgttaaacacaataacctcgctcctcgccgggtctgttccagtcggg 60
Db - 3922 ATGTCAAATCTTACGCTTAACAGCAATACCTCGTCTCGCCGGGTCTGTTCAGTCCCGG 3863

Qy 61 ggggacaacgggctgggtggtcataatgcaaatctcgcttgggggcaacacccatcgat 120
Db 3862 GGGGACAACGGGCTTGGTGTCTAATGCAAAATCTGCTTGGGGCAACACCCATCGAT 3803

Qy 121 cggcaaacattgagcaaatggctcaattattggtgggaactgtttaagctcagtcgtcg 180
Db 3802 CGGCAAAACCATTTGAGCAATGGCTCAATATTGGCGGAACCTGTTAAAGTCACTGCTATCG 3743

Qy 181 ccacaatcagggttaaTgcggcaacggagccgggtggcaatgaccagactcacaggagttggt 240
Db 3742 CCACAATCAGGTAAATGCGCAACCGGAGCGGTGGCAATGACCACTACAGGAGTTGGT 3683

Qy 241 aacgctggtgctgaacggagcaaaagggcagcagcaggaacacactccgagtcgtgacagt 300
Db 3682 AACGCTGGCGGCTCGAATGGAGCAAAAGGSCACAGCAGGAACCACTCCCGCAGTCTGACAGT 3623

Qy 301 cagaacatgctgagtgagatgggcaaacacgggtggatcgatcaggccatcacgccgatggc 360
Db 3622 CAGAACATGCTGAGTGAGATGGGCAACCAACGGGCTGGATCAGGCCATCAGGCCGATGGC 3563

Qy 361 caggcggtggcagatcggcgataatctcttactgaaagccatgctgaagcttattgca 420
Db 3562 CAGGCGCGGCGCAGATCGCGGATAATCCGTTACTGAAAGCCATGCTGAGCTTATTGCA 3503

Qy 421 cgcataTggacggccaaagcagatcagtttgcccaacctggttaoaggcaaacagtcgcc 480
Db 3502 CGCATGATGGAGCGGCAAGCGATCAGTTTGGCCAACTGGTTACGGCAACAGCAGTCC 3443

Qy 481 tcttcggtactcttcctatcgtgggttccctctttaaagatctatcagggggaagcc 540
Db 3442 TCTTCCGGTACTCTTCATCTGCGCGGTTCCTCTTTAAACGATCTATCAGGGGGGAAGGCC 3383

Qy 541 ccttcggcaactcccttcctcgggaactactctccgctcagtaacctcttcaccccccattcc 600
Db 3382 CCTTCCGGCAACTCCCTTCCTCGGCACTACTCTCCCGTCAGTACCTCTCTACCCCCATCC 3323

Qy 601 acgcaaacgtcccttaactcaccgcttgatttccctcttctcccaacaaagcagccggg 660
Db 3322 ACGCAACGTCCTTACCTACCGCTTGAATTCCTTCTCCCAACCAAGACGCGCGG 3263

Qy 661 ggcagcagccggtaaacgatcatcctgaccctgtgttagcgggagtcgagccgga 720
Db 720
```

```
Db 3262 GGCAGACGCGGTAACCGATCATCTTGACCCCTGTTGTAGCGCGGCATCGGGCCCGGA 3203
Qy 721 aattcgtggccttcacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
Db 3202 AATTCCGTGGCCTTCACAGCGCGCGCGCTAATAGACGGTGTGTCATGACACCATTAACC 3143
Qy 781 gtgaaagcgggtcaggtgttgatggcaagagacaaaccccttcacgcgcgcgcgcgcgcgcgc 840
Db 3142 GTGAAAGCGAGTCAGGTGCTTTGATGTGCAAGACAAACCTTCACGCCGCGGTTCAGAAITA 3083
Qy 841 ggcgatggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
Db 3082 GCGGATGCGCGCGCACTGTAAGAACCCAGAAACCGCTGTTTATATCTGGAAGACGGTGCACG 3023
Qy 901 ctgaaaaacgtcacaccatggcgacgacggcgcggtggtggtgagggcggtattaccgtttaagccaaac 1020
Db 2962 AAATAGACAATCTGCACGTTTACCAACGTTGGGTGAGGACGCGATTACCGTTAAGCCAAAC 2903
Qy 1021 agcggcgcaaaaaatccacgcttgaaatcactaaacgctccttcgagcagcgcctctgac 1080
Db 2902 AGCGCGGCAAAAAATCCACGCTTGAATCACTAACAGTTCTTTCGAGCAGCGCTCTGAC 2843
Qy 1081 aagatcctgcagctgaatgctgcgatacactgagcgttgacacgtgaaagcgaagcaagac 1140
Db 2842 AAGATCCTGCAGCTGAATGCGCGATACCTAACCTGACGCTTGACACGTGAAGGCCAAGAC 2783
Qy 1141 ttggtacttttgcgcacacggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1200
Db 2782 TTTGTTACTTTTGTACGCACCTAACCGCGGTCAACAGGGTAACTGGGATCTGAATCTGAGC 2723
Qy 1201 catatcagcgacgaagcagcgtaagtctcgttcgtttaaagcagatagcagggggtctaac 1260
Db 2722 CATATCAGCGCAGAGACGGTAAAGTTCTCGTTAAAGCGATAGCAGGGGCTAAAA 2663
Qy 1261 gtcaatcacagtgatctcctcgtggtggtggtggtggtggtggtggtggtggtggtggtg 1320
Db 2662 GTCAATACCACTGATATCTCACTGAGTGTGTTGAAACCACTACAAAGTCCCGATGTC 2603
Qy 1321 gccaaacctgaagggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1344
Db 2602 GCCAACTGAAGGTGGCTGAATGA 2579

RESULT 4
EAU97504/c
LOCUS
DEFINITION EAU97504 8143 bp DNA BCT 03-MAR-1998
Erwinia amylovora HrpW (hrpW) gene, partial cds; disease specific
operon, complete sequence; and lysR homolog gene, partial cds.
ACCESSION U97504
VERSION U97504.1 GI:2668586
KEYWORDS
SOURCE Erwinia amylovora.
ORGANISM Erwinia amylovora
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Erwinia.
REFERENCE 1 (bases 1 to 8143)
AUTHORS Bogdanove,A.J., Kim,J.F., Wei,Z., Kolchinsky,P., Charkowski,A.O.,
Conlin,A.K., Collmer,A. and Beer,S.V.
HOMOLOG AND FUNCTIONAL SIMILARITY OF AN HRP-LINKED PATHOGENICITY
LOCUS, dsPEF, OF ERWINIA AMYLOVORA AND THE AVIRULENCE LOCUS AVR/E OF
PSEUDOMONAS SYRINGAE PATHOVAR TOMATO
Proc. Natl. Acad. Sci. U.S.A. 95 (3), 1325-1330 (1998)
98115919
REFERENCE 2 (bases 1 to 8143)
AUTHORS Bogdanove,A.J., Kim,J.F., Wei,Z.-M., Kolchinsky,P. and Beer,S.V.
Direct Submission
TITLE Submitted (07-APR-1997) Plant Pathology, Cornell University, 334
JOURNAL Plant Science Bldg., Ithaca, NY 14853, USA
FEATURES
Location/Qualifiers
```

```
source
1. .8143
/organism="Erwinia amylovora"
/strain="Ea321"
/db_xref="taxon:552"
/note="disease specific operon"
complement(<1. .501)
/gene="hrpW"
complement(<1. .414)
/gene="hrpW"
/note="harpin"
/codon_start=1
/transl_table=11
/product="HrpW"
/protein_id="AAC04849.1"
/db_xref="GI:2668587"
/translation="MSILTLNNTSSPGLFOSGGDGLGGHNSALGQQPIDROTI
EQAOALLAELLSPQSGNAATGAGNDOTGVGNAGLNRGTAGTTPQSDSQN
MUSEMGNGLDQAITPDGGGGQIGDNPILKAWLKL"
complement(473. .501)
/gene="hrpW"
/note="hrpL-dependent promoter consensus sequence, or 'hrp
box'"
542. .570
/gene="disease specific operon"
/note="hrpL-dependent promoter consensus sequence, or 'hrp
box'"
542. .6609
/gene="disease specific operon"
542. .6128
/gene="dspe"
612. .6128
/gene="dspe"
/note="AvrE homolog; pathogenicity factor; avirulence
protein"
/codon_start=1
/transl_table=11
/product="DspE"
/protein_id="AAC04850.1"
/db_xref="GI:2668588"
/translation="MELSLGTEHKAHVTAHNPVGHCVALQQSSSSPONAASL
AEGNRGKMPRIHQPTSAADGISAHQKKFSLRGCLGTRKFSRSPQGGPQTHS
KATLDRLDARDGDEHAAAPDAARLTSGGVKRRNMDDMAGRPVGGSGEDVP
TQQRHLNFMQOMTQMLSKMAHPASNAGDLQHSPPHIPGSHHEKEEPPVSTK
ATTAAHADRVEIAQEDDSEFOOLHOARLARENPPQPLGVATPISARFOPKLTAV
ASVLEGDTTQSPKPSMLKSGAGVTPLAATLDKGLQALPNPALNTLLKQTL
GRDQHYLAHSSDGSQHLLDNKGHLEDIKSTATSYSLNHPGEIKKLAQGT
GSVVDGSGKTSLSGSGTQSHNKMLSQPGEAHRSLLTGIWOHPAGAHPQSESLRH
DDKIHILPELGVMSADKTHSQRGDKLYALKDNRTLONLSDNKSKSLVDKI
KSYVDQRQVAILDTDFGRHKMSIMPSLIDSPESHISLSLHFAHAHQGLLHKSELE
AQSAVISHGLRVVADSEGLFSAAPKQGDGNELKAMPQHALDEHFGHDHIGSEFF
HDDHQLNALVKNFRQOHACPLGNDHOFHPCGNLTDLVLDNQLGLHNPPEHIL
DMHGLSLALQSGKLYHFDLTKGWTGAESDCKLKGDLGAAYLLKQGEVYRLNQ
STSSIKHGTENVFSLPHVRNKEPEPDALQGLKDKQAQMAVIGVYKYLALTEGDIR
SFQIRPQQLERPAQTLSREGISGELDIHVHDKONLYALTHEGEVHFHPREAWNG
AESSMHWKLLAPQSESKLSLDMSEHKPIATFEDGSQHLKAGWHAYAAPERGPLA
VGTSGVTNFRLMQGVKGVIPGSGELTVKLSAQGTGMTAGRKVSVSKFSERIRAYA
FNPTMTPRIPKAAAYATOHQWQREGTLPYEMOGALIKOLDADHNRHNAPOPDLOS
KLETLDLGHGAEALLDMKFRPDELEQSLATRSVTVLGOHGVLKNGEINSEKPSPG
KALVQSFVNRSGDLSKSLQAAVHATPPSAESKLSQMLGHFVSAGVDMSHQKGEIPL
GQRDPDKTALTUKRLDITVTIGELHELADKALVSDHKPDADQIKOLRQOQFDTLR
EKRYFENPVKHYTDGFTHTNKALEANYDAVKAFINAFKEHGVNLTTRVLESQSGA
ELAKKLTNLLSDGSMSFSRSGVSTVFVPTLSKKVPVPVIPAGITLDRAYN
LFSRTSGGLNVYFGRDGGVSGNIMVATGHDVMTGKTSAGNASDWSLAKHKISP
DLRIGAAVSTGLQGTQLNSLKFLEDELPGETHGLTHTLTPAELLQKIEHOMKQ
SKLTSVSDTSANFLDLRAGINLNEGSKPNGVARTVSAGLSASLANAASRSTSSGQ
FGSTTSASNRPFTLVNGVAGANLTAALGVAHSSSTHEGKPVGIFPAFTSTVSAALAL
DNRTSQSLKLEKRAEPTVNDISLTSLTGKHFKDSATTKMLAALKLEDDAKPAEQL
HILQHFSAKLVGDERVEAVRNKLVIROQAADSHMELGSASHSTYNNLSRINN
DGIVELLHKKHFDAAIPASSAKRLGEMMNDPALKDIKLOLQSTPFSSASVSWELDGL
RQOTEKAILDGKVGREEVGLFDQRNRLRVKSVSVSQSKSEGTNPALLGLGTSNA
ANSMERNGTGTFNKYIQDQONTFRFTLEGGIAQANPQVNASULTDLKLEGLMKS"
6190. .6609
/gene="dspf"
```

```
CDS
6190. .6609
/gene="dspf"
/codon_start=1
/transl_table=11
/product="Dspf"
/protein_id="AAC04851.1"
/db_xref="GI:2668589"
/translation="MTSSOORVERFLQYFSAGCKTPIHLLKDGVCALYNBQDEAAVLE
VPOHSDSLLLHCRITTEADPQTSITLYSMLLQINFMAARGCMLDELHLYRLCTFOQ
SLEHLDEASFSFSDIVSGFFTEHAAREVREYIAQLDESSAA"
7361. .78143
/note="partial ORF"
/codon_start=1
/transl_table=11
/product="lysR homolog"
/protein_id="AAC04852.1"
/db_xref="GI:2668590"
/translation="MFISKMQVFTLYOERSLKTAADKLCTVPPVSRMLKLTGEGV
GEQLFIERNNTPTPAADCIYQQLPLYMALKNMTRKHPEQOEFLIASPYISTISLS
DLTITSESAIPKCSSTKYAECIHPDDDIYISLPVECPMHFPIRMDLVLECCSKDV
GDKWLHMPITLAESETKKLYFORSLAEHLARGSGVLRQVDNPSHLQSTFKSGEGLLF
RLPPKNMVKDYQSLPVISSHCLPIYVNTLKKDKRHNCETAKL"
BASE COUNT 2181 a 2027 c 2177 g 1758 t
ORIGIN

Query Match 30.8%; Score 414; DB 2; Length 8143;
Best Local Similarity 100.0%; Pred. No. 7.1e-95;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtcaattcttcgcttaacaacaatacctcgtcctcgcgggtctgttccagtcggg 60
Db 414 ATGTCAAATCTTACGCTTAACAACAATACCTCGTCTCGCGGGTCTGTTCCAGTCCGG 355

QY 61 ggggacaacgggcttggtggtcataatgcgaattctgcgttgggggcaacaacccatcgat 120
Db 354 GGGGACACGGGGCTTGCTGTGTCATAATGCAAAATCTCGCTTGGGGCAACAACCCATCGAT 295

QY 121 cggcaaacatttgagcaaatggctcaattattgacgaaactgttaaaagtcactgctatcg 180
Db 294 CGGCAAAACCAATTAGGCAAAATGGCTCAATATTGCGGGAAGCTGTTAAAGTCACCTGCTATCG 235

QY 181 ccacaatcaggtaagtgcggcaacccgagcgggtggcgaatgacagactacagagtgctggt 240
Db 234 CCACAATCAGGTAATCGGCAACCCGGAGCGGGTGGCAANTGACCACTACAGGAGTTGGT 175

QY 241 aacgctggcggcctgaacgagcaagaaaggcacagcagcaggaaccaactccgagtcgacagt 300
Db 174 AACGCTGGCGGCTGAACGGACGAAAGGCACAGCAGGAACCACTCCGCAGTCTGACAGT 115

QY 301 cagaacatcgtgagtgagatgggcaacaacgggctgagtcagcagccatcagcccgatggc 360
Db 114 CAGAATCGTGTAGTGAGATGGGCAACACAGGGCTGGATCAGGCCATCAGCCCGGATGGC 55

QY 361 caggcggcgggagatcgcgcgataatcctttactgaaagccatgctgaagctt 414
Db 54 CAGGCGCGGGGAGATCGCGGATGATCCTTTACTGAAAGCCATGCTGAAGCTT 1

RESULT 5
AF037983
LOCUS AF037983 1833 bp DNA BCT 05-OCT-1998
DEFINITION Pseudomonas syringae hrp type III secreted protein (hrpW) gene,
complete cds.
ACCESSION AF037983
VERSION AF037983.1 GI:3695000
SOURCE Pseudomonas syringae pv. syringae.
ORGANISM Pseudomonas syringae pv. syringae
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 1833)
AUTHORS Charkowski,A.O., Alfano,J.R., Preston,G., Yuan,J., He,S.Y. and
```

QY	1108	aacctgagcgttgacaaagtgaaagccaaagacttttggtaactttttgtacgcactaaagcgc	1167
Db	1354	CACCTGAAGTCGACGGTGTTCAAAGCGCCACCGATTTTCGGTACGCTGGTCCGACCAATGGT	1413
QY	1168	ggitcaacag--ggtaactaggatctgaatctgaacatcatcagcgaagaagcggttaag	1224
Db	1414	GGCAAGCAGTTTGATGACATGAGTGTGAGCTCTAATGGCGTGGACGCGACCCATGGCAAG	1473
QY	1225	tctcgtctgcttaaaagcagatagcgagggggtcaaacgtcaataaccagtgatctcactg	1284
Db	1474	TTGCGACTCGTCAAAAACGACAGTGCAGATCTGAACACTGGCCACTGGCGACATCGCCATG	1533
QY	1285	ggtagtggttgaaacacactacaaagtgcagatgctccgcccaacctgaa	1331
Db	1534	ACCGACGTCGAAGCAGCGCTTGATGACCAAGCAAGCGCTCTACCCAGCA	1580

RESULT	6				
AF005221	AF005221	1729 bp	DNA	BCT	05-OCT-1998
LOCUS	Pseudomonas syringae hrp-secreted protein HrpW (hrpW) gene, complete cds.				
ACCESSION	AF005221				
VERSION	AF005221.1				
KEYWORDS	GI:3694998				
SOURCE	Pseudomonas syringae.				
ORGANISM	Pseudomonas syringae Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.				
REFERENCE	1 (bases 1 to 1729)				
AUTHORS	Charkowski,A.O., Alfano,J.R., Preston,G., Yuan,J., He,S.Y. and Collmer,A.				
TITLE	The Pseudomonas syringae pv. tomato HrpW protein has domains similar to harpins and pectate lyases and can elicit the plant hypersensitive response and bind to pectate				
JOURNAL	J. Bacteriol. 180 (19), 5211-5217 (1998)				
MEDLINE	98422476				
REFERENCE	2 (bases 1 to 1729)				
AUTHORS	Charkowski,A.O., Conlin,A.K. and Collmer,A.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-MAY-1997) Plant Pathology, Cornell University, Ithaca, NY 14853, USA				
FEATURES	Location/Qualifiers				
source	1..1729				
gene	/organism="Pseudomonas syringae"				
promoter	/strain="DC3000"				
	/db_xref="taxon:317"				
	/note="pathovar tomato"				
	pathovar: Lycopersicon esculentum"				
	268..1693				
	/gene="hrpW"				
	268..292				
	/gene="hrpW"				
	/note="hrp box"				
	406..409				
	/gene="hrpW"				
	419..1693				
	/gene="hrpW"				
	/note="harpin; hrp-secreted protein"				
	/codon_start=1				
	/transl_table=11				
	/product="HrpW"				
	/protein_id="AAC62526.1"				
	/db_xref="GI:3694998"				
	/translation="MSGITPRPQTTPLDFSALSGKSPQNPTEQNTQQAIDPSA				
	LFLGSDQDQKVFNGTPTDSTVQNPDSKPNDSQNIAKLISALIMSLQMLTNSNKKQ				
	DTNQEOPDSQAPQNNNGILGTPSADSGGDTPTATGGGGDTPTATGGGGGDTPTATG				
	GGSGGGGTPTATGGSGGTPTATGGGGGYFTQITPOLANPRTSGTGSVSTAGST				
	EQAGKINIVYDKTIVKGAQVEFDGHTAFADKSMGNDGQENOKPWFELAGATLKAV				
	ENGLNEVDGDIHYKAKNAQVETIDNHAQNVGDLITVKGEGGAIVNLNITKNSAKGA				
	DQVKLVNANTHLKIDNFKADDFGTWRTNGNSQVETIDNHAQNVGDLITVKGEGGAIVNLNITKNSAKGA				
	SDKLVIATGNIAMTIDVKAHYDKTQASTQHTEL"				

[illegible]

```
FT PWTDEAVANATRFSPDSHLQSGDGRISMLATDPDQSPSSGSKIGDSGPIPPPREP
FT MLWNSGRFELKDEKLVNRPSEPGSQSLDAKGRPDFSTENPTGLAPLLDLSILATPKQT
FT YLWHSQDKVGHQLOAAGHFLHLAODSSLAIVRSNEALLIEGKPPKPAVKMEREDG
FT NIHTDASGRKTQPLGKAHTJTNVLVSHDGRMRVHEDRLYQFDP ISTRWKIPEGI
FT EDHFNLSITGSGNSVYAKSDADVDDVSSPMPHVEVEDLOSFSVADNRALLSGKTT
FT QAILLTHSPVIGLTKPKTKGLEDDGGKAQAAVGLSGDKLFIADTQGRUISADRSAF
FT EGGDPKLLKMPENQANFQELVPLGGHNKVTGFINGGDGGVHALIKNROGETHSLADEQ
FT SSKLQSGWNLNVALVNNRGLTWPPPTTAADRLNDRAGLVSEGRIOEMWATPBCW
FT KEAGIKOIDRLQGADSNAYVLKGGKHALKIAAEHPNMAFDRNTALTAQKRVEMG
FT DEIGLDDRVTKAFAMYSNKREVALDDQNKLTASHKDKHPTVLDI PGLEGGIKSLIDE
FT KHNHLATSTGGLYCLPKEAWQSKLGDOLRARTPTVALPGQGVKALFTNDDNLSAQ
FT IEDABGKLMQAKQWQRFQRPVEENGLNDRHSRTNKTWRIPKTLGLTRMDVTM
FT FGRSGVEKSKASTSEFIRANYKNTAETPRMKNVGDHIQHRYGQRLGLKEVYTESM
FT LKQLELTHESGGRPPARGQDLKARITALEAKLPGQATLVKELETLRDELENSHYTAL
FT MSIQSGYKAKNLKQOQILNQHGELAKPSVMOFGKRLADLGT KLNPKSGSHDLVKEL
FT QDALTOVAPSNPTKLLGTLKHQGLKLSHOKADIPLGORRDASEDHGLSKARLADL
FT VTLKSGALLDQVQLPOSDIEPLEQKLTATLDVTVYGENPVKVYTDGFTDNKALEG
FT YESVYTELKSKADHAVSNVNRATSGKQDAELAGFKSKMLKOLEHGDDEVLQRSYG
FT VNLTPFIIADKATGLMPTAGATGNRNYILNAERCBBGVLYLIDISEGAVNSGGFAG
FT KQWPGFDANPNARSVDNNRNLTPNFRLGVDVATVAASORAGVNVFVNPDBDIDAF
FT VDDLEFQOLNPQLKKAHVHESYEARENFEDLTAGGTADIRAGINTLEDPRDNAPNS
FT DSFSAVVRGGPAAINTVNLMTYTOYSLTKNDKTELKEGKNRFRFLNNVTAGGOLRAQ
FT IGGSHPTAGTPPASAGPTPASOTAAANLGGALNFSVENRTVKRIKRYNVAKPIITTEG
FT LKLSKGLGEAPLNTKAKLAEADPLNARYTGCKKDEVDLQAGLDLEELFADIPPPK
FT DNKOYKALRODKRAAVHRASANKHVSMDNAREFTSKTNLSGUSSEITLKNSSVRD
FT ASFGNATRVAEFRQDPKRLAMLEKEMEGSIGTLARVLEPKDLSLVKIDEGSLUNGMT
FT QDLSLMELEDRNEMRIKRLVVFHTAQENFTSPTPLVSYNSGANVSVTTLGLRINEVY
FT GADQKPIGYTFDGELSRPSASLKEAAGDLKKEGFELKS"
FT complement(8226..8621)
FT /codon_start=1
FT /db_xref="SPTREMBL:Q9JP37"
FT /note="ORF2"
FT /transl_table=11
FT /product="unknown"
FT /protein_id="AAF71500.1"
FT /translation="*WTKTSQDPARFINSIGAOLGTLTLONGVCALYDQGNNEAII
FT ELPHSEMVIFHCIRKCPRAPDRLRLLSLNFVARLHGCWFVADQGDVRLCAQRELA
FT SLDPAPCDVTRGTFSQAREARFLQA"
FT complement(8960..10831)
FT /codon_start=1
FT /db_xref="SPTREMBL:Q9JP36"
FT /note="ORF3"
FT /transl_table=11
FT /product="unknown"
FT /protein_id="AAF71501.1"
FT /translation="*MISRIIGGAGGVKLSRVNQOHDVTPAQTAHPNNAVTAAGMNPPLTPD
FT QSGSHAYTESSAGARLNVAARHTQLLQAFKAHEGTAPVSGAPMISRAALLIGSLLOA
FT EPLPEYMAEKLSPERVLQKQFQSDLDQORLEKFAEQPGQIPDKAEVQLIKGPAQSVAD
FT QLEHFQMLHDSAPATVGOHAKADKATLAVSQTALGEYAGRSKAKAIGBGLSNSIASLTDH
FT ISALDLTLDAAEQNKESLHADROLVDKTLVLGLHADFYKSPKARLXASVAHTQLD
FT NVVSDLVARTNTVGGWKAGPIVAAAVPQFLSSMTHLGYRLSTSDKLRDITIPETSSDA
FT NMLKASIGTVAGTAEHTVNSVVKPMFQAALOKTGLNRLNVMVPMKAVDNTNTPDPPE
FT LKSHGELVKTPEEVAODKAFVKSERALLNOKVQGSSTHPVGEIMAYSAFCGSOAVR
FT QMLNDVQINGQTLASALASGFCGAVSASQTLQLKSNVDPQGRKKIPVFTPDRAES
FT DMLKDLKGMDLRFPSTVNTSYTKALSGIOSSELTALPPVTVAQAEAGASGTLISAGAILR
FT NMLAATGYSLLTYTNSVTAFAKALKAAAGMGATPMLDRTEIL"
FT complement(110828..111322)
FT /codon_start=1
FT /db_xref="SPTREMBL:Q9JP35"
FT /note="ORF4"
FT /transl_table=11
FT /product="unknown"
FT /protein_id="AAF71502.1"
FT /translation="*MTNNDQVHTLI NEICALSLISTPERFVESAENFKISEVDFTLQFOD
FT RDEGRAVLIVGDMCALPARGESALLALMDINFMHFAGASHSPAFSNAQTRGVLLMGSV
FT ALERASAGVLLLMKSFSDIAKKEWREHFGMGQATAGSSSTDOQPVAPAAKRESLSAPORF
FT Q"
FT complement(12902)
FT /codon_start=1
FT /db_xref="SPTREMBL:087264"
FT /transl_table=11
FT /gene="hrpw"
FT /product="hrpw"
```

```
FT /protein_id="AAF71503.1"
FT /translation="*MSITGTPRPQQTTPDLDFALSLSGKSPQPMTFGQNTQQAIDPSAL
FT LFGSDTKQDNVFGTDPSTVONPDASPNDSQNSIAKLISALIMSLQMLTNSNKKQDT
FT NOEQPDSQAPQPNNGGLGTPSADSGGGGTPTATGGGGGDTPSATGGGGGDTPTATGGG
FT SGGGCTPTATGGSGGGTPTATGGEGGVTPQITPOLANPRTSGTGSYSVDTAGSTEAG
FT KINVKDITIKVAGEVEFGHATFTADKSMNGDQGENQKPMFELASGATLKNNVLNG
FT EVGDIHVYKANQAQEVITDNNHAQNVGDELITVKGEGAAVTNLNKINRSKAGGADDKVQ
FT LNANTHLKIDNFKADDEFTVTRTNGGQFDDMSIELNGIEANHGFALVKRSDSDDLKA
FT TGNFATDVHAYDKTQASTQHTEL"
FT complement(12986..14446)
FT /codon_start=1
FT /db_xref="SPTREMBL:Q9JP34"
FT /note="ORF5"
FT /transl_table=11
FT /product="unknown"
FT /protein_id="AAF71504.1"
FT /translation="*MHNRRVQOPVTATDSFTASDASLASSSVRSVSDQOREINAI
FT ADYLTHVFAHKLPPADSADQNAVVDVHNAQITALTIETRSLRHIFGEPTPATIADTFA
FT KAEKLDRLMTTSGALRATPFAMASLDQYQWPAINKGDMLPAPLKPITPLISGALSAM
FT DQVTKRMDRATGDLHYLSASPDRLHMAAASVKRHSPLARQVLDTVAVQTVSARNA
FT VRTVLPALASRAPVAGVDLGVSGMGLAANAGFGNRLLSVQSRDHQRGALVGLKD
FT KEPKAQLSEENDWLEAVKAIKSASYSCAALNAGRMAGLPLDMATDAMGAVRSLSVSS
FT LTONGLALAGSFAGVGLQEMATKNITDPATKAVSOLTNLAGSAAVFAGMTTAAALTDD
FT PAKKASRFQDITVKSTASSTTGVADQTVKLATVYKDMGGEATHTTGASLRWTVNRLR
FT QPKAREADIEBEGTAASPSPIPRPMRS"
FT complement(14818..15891)
FT /codon_start=1
FT /db_xref="SPTREMBL:Q9JP33"
FT /note="ORF6"
FT /transl_table=11
FT /product="unknown"
FT /protein_id="AAF71505.1"
```

Query Match 11.5%; Score 155.2; DB 16; Length 20052;

Best Local Similarity 57.18; Pred. No. 6.4e-29;

Matches 330; Conservative 0; Mismatches 233; Indels 15; Gaps 2;

```
Qy 745 ggcgcgaatcagcgggtctgctgacaccattaccgtgaaagcgggtcaggtgtttgat 804
Db 12288 GCCGCAAGATCAATGTGTGTAAGACACACCATCAAGGTGCGGCTGGCGAAAGTCTTTCAC 12347
Qy 805 ggcgaagagaaacacctcaacgcgggttcagaaataggcgtgagcgcagctcgaaaac 864
Db 12348 GCCACGCGCGCAACCTTCACATGCCGACAAATCTATGGGTACGAGACACGCGGCGGCAAAAT 12407
Qy 865 cagaaaccgtgtttatctactgaaagcgggtccagcctgaaaaacgtcacctatggcgac 924
Db 12408 CAGAAGCCCATGTTGCGACCTGGCTGAAGCGCTACGTTGAAGAATGTGAACCTGGGTGAG 12467
Qy 925 gacggggcgagtggtattctcttttaagc-----gtgatgcaaaaatagacaaat 972
Db 12468 AACGAGGTGCGATGCGATCCACGTGAAGCCAAAGCGCTCAGGAAGTCAACCATTTGACAAAC 12527
Qy 973 ctgcagctcaccaacagctggtgagagcggattaccgttaagcgaacacagcgcgggcaaa 1032
Db 12528 GTGCATGCCCAAGAACCTGGTGGTGAAGACCTGTATACGGTCAAAAGCGGAGGCGGACGCG 12587
Qy 1033 aaatccacgcttgaatacaactaacagttcccttcagcagcctctcgacaagatctctgag 1092
Db 12588 GTCACATACTGACATCAAGAACAGACAGTCCCAAGAGTGTCAGACGACAGGTTGTCGAG 12647
Qy 1093 ctgaatgcgcgactactaacctgagcgttgacaacgtgaaagcgaacgacttttgactttt 1152
Db 12648 CTCAAAGCGCAACACTCACTTCAAAAATCGACAACITTCAGAGCCGACGATTTCGCGACGATG 12707
Qy 1153 gtacgactaacgcgcgttcaacag---ggtaactggggtatctgaaatctgagccatcatcagc 1209
Db 12708 GTTCGCAACCAACGGTGGCAACAGCTTTTGATGACATGAGACATCGAGCTCGAAGCGGATCGAA 12767
Qy 1210 gcagaagacggttaagtctctcgtttaaaagcgatagcgagggcggttaaacgtcaatacc 1269
Db 12768 GCTAACCAACGCAAGTTCGCCCTCGTGAAGAACGACACAGTACGACGTCTGACAGCTGCGCAACG 12827
```



```
/protein_id="AAC67261.1"
/db_xref="GI:2618832"
/translation="MKSLSLTIGFALSLLAAGSNDAKEETKDTGSKTEATASEG
BELVQOSVCGCHKDLEGVGNLQEVGKIDEHKIESIKNRGRNMPKGLVDDNEAA
VIAWLSEKK"
RBS      2806..2813
          /gene="ftsE"
          /gene="ftsE"
          /gene="ftsE"
          /gene="ftsE"
          /codon_start=1
          /transl_table=11
          /product="cell division ATP-binding protein"
          /protein_id="AAC67262.1"
          /db_xref="GI:2618833"
          /translation="MIENKEYKAYPNCVKALNGISVTIHPGEFVVVPGSGAKSTF
          IKMIYREKPTKGLINLHKDLATIKKEIIPFVRKKIGVVFQDFKLPKLPKLVFNFAF
          ALEVGQPSVKKRVLVLELDLQVKKHARQFPDQSGEQQRYSIARSIVNPDVVI
          ADEPTGNLDPDTSWEVMTLEEINNRGTTVVVMATHNKIEIVNTMKRRIAIEDGIIVR
          ESREYGSYD"
          gene      3476..4385
          RBS      3476..3484
          /gene="ftsX"
          /gene="ftsX"
          /gene="ftsX"
          /gene="ftsX"
          /codon_start=1
          /transl_table=11
          /product="cell division protein"
          /protein_id="AAC67264.1"
          /db_xref="GI:2618835"
          /translation="MIKILGRHLRESFKSLGRNTWTFASAVTVLILVGVFLVIM
          LNLNMTAENKQVEIKVLIDLTDADQAKDLNDIKELKGIQSVTFSSKPELDOLV
          DSFGDSCKSLTMKQENPLNDAFVVKTDPHDTPNVAKKIEKMDHVYKVTYKGEYSR
          LFKVYGRNIGIALIIGLVTFAMFLISNTIKITIFARRKEIEIMKLVGATNFIWRP
          FFLBGLLGVFGSVIPIALVSTYYVIGVWVPKVGQSVSLPYNPVPVQVSLVLA
          IGAVIGWGSLSITSIRKFLRV"
          RBS      4511..4516
          /gene="yvJb"
          /gene="yvJb"
          /gene="yvJb"
          /gene="yvJb"
          /codon_start=1
          /transl_table=11
          /product="putative protease"
          /protein_id="AAC67263.1"
          /db_xref="GI:2618834"
          /translation="MNQKIMAVIAAGSMLFGAGVYAGINLLEMDKPTAAVPATAQA
          DSRDKAMDTEKAYELISNEYVEKVOREKLEGAIOGMLSTLNDPYSVYMDKOTAKQ
          FSDLSDSFEGIGAEVGMEDGKIIIVSPFKKPAEKAGLKPNDIEIISINGESMAGKDL
          NHAVLKTRGKGSVSMKIORPQTKKOLSPRIKRAEIPLETVEASEKKVOGSHVGYIA
          ISTSEHTAEDPAKALRELEKEIEGLVIDVRGNPGGYLQSVBEILKHFVTKDQPYIQ
          IAERNGDKRTFSLTHKKATPVNVIDKGSASEILAGALKAEAGHYDVVGDTSFGK
          GTQVQAVPMGDSNKKLTLTKWLPNGNWIHKGIEPTIAIKQPDYFSAGPLQKEPL
          EYLMNEDVKHAQVLLKGLSPDQREDGYFSKMKKAMVAFQDNKLNKTKGVIDTRTA
          6773..7977
          /gene="yvJd"
          /gene="yvJd"
          /gene="yvJd"
          /gene="yvJd"
          /note="transmembrane protein"
          /codon_start=1
          /transl_table=11
          /product="yvJd"
          /protein_id="AAC67265.1"
          /db_xref="GI:2618836"
          /translation="MSVQMGTELLKSAGLFLHPLFWFFIIITLAFGVYRKREKTF
          HTRADIYDCLKFYTKGLIPGLLSLVLGGLGISIPGLLAIITAVITAAAFILRAN
          WMSAYIVSVSMLIGLGIQIQOAEFFLERPQGFVAVWPAVAVFLGLLIITIEGAVAYR
```

```
SAHVRTSPALVSSRGLPQIQQOLANRVWLLPLFLVPGNGLESLSLWMPVFTVPGGSF
HPLATPYFVGFGORVQVGLPETSIRITAKRVCIILGLAVAVLGAASLLWTPLAGAAVCT
ALLGFIPLSIKORVNDNAPYFSKRDGLAVGLCIIPNTPAEDLELKGTEITKNGI
PVKNVSDYFALQHNRAIVKLEIIIGLNGEIRFQKRASTEGEHGELGLFVKKDDREDEA
VASGS"
complement(8025..9373)
/gene="yvka"
complement(8025..9359)
/gene="yvka"
/note="belongs to the major facilitator family"
/codon_start=1
/transl_table=11
/product="yvka"
/db_xref="GI:2618837"
/translation="MSSRKWALIVSLLGAILVPINSTMIAVALSSISHTYNBSIAS
ITWVTVLIYMAVTQIAGLGMNGKMTYLMGVLGLFIASLGCALSPSLLLIVF
RALQAVGALLTPNSIAIRHVSEKRLPKVFGFGLGAGLGAALPGFISLIDFSF
WHSIFWNPFLAIALFTALTMFPQYKENSADPLDITGSLLAGSIVSILITKNEA
PWGTVTVSVLILFPLFRREKTOHPIDIFALFKSFTTNANLSVLNLMYAVL
LIMPLEWNOFGLATNSGMALSVFSIFMSASNMVGAOLHHKWKAKKTIIFLSEFAMAG
ANLLFLLSSSHSVLEMLSLILGLASGVGLTSMQVSSLATVDPGMSGVASGIFSTF
RYFSIISALIGLISGTHLFLMILFAVSIILGVFVSGIKSDETARIEKNSA"
complement(9366..9373)
/gene="yvka"
complement(9390..9972)
/gene="yvkb"
/note="9959"
complement(9390..9959)
/gene="yvkb"
/note="belongs to the TetR/AcrR family of transcriptional
regulators"
/transl_table=11
/product="yvkb"
Query Match      6.8%; Score 91; DB 1; Length 47739;
Best Local Similarity 58.8%; Pred. No. 1.4e-12;
Matches 157; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
QY 748 gctaatacagagcgtgctgcatacacaccattaccgtagaagcgggtcagggtgtttgatgac 807
    || || || || || || || || || || || || || || || || || || || || || ||
DB 37642 GCAGCTGACAAAGTGTGTCACGAAACAATTATCGTACCAAAAAAATACAAACATATCACGGG 37583
QY 808 aaagagcaaaacccaccgcgcgttcagaattagcgatgcgcagtcctgaacacacag 867
    || || || || || || || || || || || || || || || || || || || || || ||
DB 37582 AAAGACACACCGGTTTGTGGCAGGGAAGAAATTAGGTGACGGAAGCCAGTCACAAACCAA 37523
QY 868 aaaccgctgtttatactggaagacggtgccagcctgaaaaacgctcaccatggcgacgac 927
    || || || || || || || || || || || || || || || || || || || || || ||
DB 37522 GAQCCCTGTTTTCGTGTGGAGGATGGAGCAACCCCTGAAAAATGTGCTGTTGTCACCT 37463
QY 928 ggggcgagtggtattcatctttacagtgatgccaaaaatagacaaatctgcacgtcaccaac 987
    || || || || || || || || || || || || || || || || || || || || || ||
DB 37462 GCAGCTGATGGCGTGCACACTTATGGAACGTTAACTTACGAATGTGAATGGGAAGAT 37403
QY 988 gtgggtgagagcgcgattaccgattaag 1014
    || || || || || || || || || || || || || || || || || || || || || ||
DB 37402 GTTGTGAGGATGCGCTTAACGCTGAAG 37376
RESULT 11
BSUB0018      BSUB0018      209510 bp      DNA      BCT      26-NOV-1997
LOCUS      Bacillus subtilis complete genome (section 18 of 21): from 3399551
DEFINITION      to 3609060.
ACCESSION      299121 AL009126
VERSION      299121.1 GI:2635827
KEYWORDS
SOURCE      Bacillus subtilis.
ORGANISM      Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
```

REFERENCE
AUTHORS

1 (bases 1 to 209510)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
Daniel, R.A., Deniot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrarri, E.,
Faulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Gollightly, E.J.,
Grandi, G., Guiseppi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullio, M.F.,
Icay, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaerstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinols, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P., Mitsu,
Mizuno, M., Moesti, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portetle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolto, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B.,
Sorokin, A., Taccini, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H., Zumstein, E., Yoshikawa, H. and
Danchin, A.

TITLE
The complete genome sequence of the gram-positive bacterium

JOURNAL
Nature 390 (6657), 249-256 (1997)

MEDLINE
98044033

REFERENCE
AUTHORS

2 (bases 1 to 209510)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Submitted (18-NOV-1997)
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES

source
Location/Qualifiers
1. .209510
/organism="Bacillus subtilis"
/strain="168"

gene
/db_xref="taxon:1423"

terminator
complement(4. .626)
/gene="yvyk"

CDS
complement(4. .29)
/gene="yvyk"

complement(45. .626)
/gene="yvyk"

/function="unknown"

/note="similar to hypothetical proteins"

/codon_start=1

/transl_table=11

/protein_id="CAB15320.1"

/db_xref="GI:2635828"

/db_xref="SPTREMBL:O34899"

/translation="MKLYTKGDKGTGLVGGRTKDKSLRVESYGTIDELNSFIGLAL
AELSGOGPEDTAELLTOHLEFGDGLAIYTERKDYKLTESYFLETRIDAVTA
EAPLKKFLLPGSKCASHLIARTITRRARRVVALMKSEETHTVLYLNRLSDYF
FACARVYNARSGIGDVEYERSAIVFRDRNSSSES"

gene
complement(649. .1977)

CDS
complement(649. .1977)
/gene="yvyrA"

/gene="yvyrA"

/function="unknown"

/note="similar to iron transport system"

/codon_start=1

/transl_table=11

/protein_id="CAB15321.1"
/db_xref="GI:2635829"
/db_xref="SPTREMBL:O34631"
/translation="MKAEGLSGGDSRLNNVSLVVEKGEFLGILCPNGSGKTYLLH
LTLGTLPKAGRVYLAGKLADYKPKELAMVLPKMQQAAETFTVEETVAGRYPF
QTGLFQOQTEKGAIVQEAQMEQTVADFAQKPIRELSGGQQRVYLAQALAQPRILF
LDPTFLDLAYQKDLDLIKRLTRESGLAAVSFVFDLNTASLYCDGLMPKNGTAGP
KQPEAVTEQSIKAVYDVTALVHQSSPKMIVIOPEKDSVKRQSIPEFALLQAGR
DILLLEIPLRTSPICAGFSWSRLTHKRLPOQDPDIEGLTACLSSGQLOQT
CMASSERLDRFVRYVEDGELSVEFCVQTFGSWILLINGCYAADOFFIKALMAEAEER
TKVLGGGTGDLIAATQQTQSENIEQRNLKKGTAECIKAEAELEF"

gene
complement(1977. .3038)
/gene="yvrB"

CDS
complement(1977. .3038)
/gene="yvrB"
/function="unknown"
/note="similar to iron permease"
/codon_start=1
/transl_table=11
/protein_id="CAB15322.1"
/db_xref="GI:2635830"
/db_xref="SPTREMBL:O34451"
/translation="MKRAFTRIRLRNRYAAYIIGFSFLAVSLIGSCSLHIPIPA
VFRVHQGGSGSDPMYTNIMNIRLPVYLAALVGAALSIAGAATFOGLKKNPL
ADPTLGVSSGASVAVTFLGLHLPVIGFTLPVLSVAALATMAAATVFFRLVHA
SMVSSTLITGVITNSFLGAFISLIITALTGNLLPIVHMLGVSVMRGVSIVILFLPF
FLGTLTLLINGRELNMVYGEDKAKILGVSVQORKMWILLIAGSLLTGSAVAVSGTIG
FVGLVPHITRLNGDTHRHLLPLSALLGAGFLVADLLSRTIIEPIELPIGITSIA
GAPVALILIRHQHGGHSL"

gene
complement(3001. .3945)
/gene="yvrC"

CDS
complement(3001. .3945)
/gene="yvrC"
/function="unknown"
/note="similar to iron-binding protein"
/codon_start=1
/transl_table=11
/protein_id="CAB15323.1"
/db_xref="GI:2635831"
/db_xref="SPTREMBL:O34805"
/translation="MKRAGIAWALLLAAYVLAGGPNADQKSKAKQKTEVPVITD
DASNDQVTIKKEPKIVSLSPNTEITAYALGDKGVVGTNTNDTYPEVKKVEKVGDM
NVNVEKVISLPLDLVAHSSMSASDAIKQLDAGITVLTVNDAAQSFSEVYSKIEMI
GEAGAEKADQLVKSHKSDLKIOEKAKTISKDEKSVFIEYSPDPDITVTGKDTFM
KHMNVYDVPDLVTRSGPRLIEGVLEAESIYDPDFKE"

gene
4343. .5165
/gene="yvrD"

CDS
4343. .5134
/gene="yvrD"
/function="unknown"
/note="similar to ketoacyl-carrier protein reductase"
/codon_start=1
/transl_table=11
/protein_id="CAB15324.1"
/db_xref="GI:2635832"
/db_xref="SPTREMBL:O34782"
/translation="MDLQREKLVLTGTSIGTGKAAKSFLOEGAIVYNGRQETV
DRTIELSGYGVHGAADLSKDEAAAFIEKVNIEGDDIDLNNLGGFFVEKDFADVT
DEENQYFEYVNSAVTRSHFLPKLAKNSGRILNIASEAGVKPLTPMIPYSMTKTA
LISLSRGMAEMTKGTVNTVNSVLPGLPTWEGVASVMEGAAGAAGQD'TDTIKDYKVN
EPTSLQRYATAEAVANTIVFLASDAASAINGTAQRVEGGIIRSL"

terminator
5145. .5165
/gene="yvrD"

gene
complement(5172. .6050)
/gene="yvrE"

CDS
complement(5172. .6050)
/gene="yvrE"
/function="unknown"
/note="similar to RNA polymerase"
/codon_start=1
/transl_table=11
/protein_id="CAB15325.1"

polyA_signal	865..870	/gene="pela"
polyA_site	884	/gene="pela"
BASE COUNT	204 a ..242 c	233 g 205 t
ORIGIN		
Query Match	4.6%; Score 62.2; DB 14;	Length 884;
Best Local Similarity	52.0%; Pred. No. 3.5e-05;	
Matches 173; Conservative	0; Mismatches 148; Indels 12; Gaps 1;	
Qy	854 agtggtaaaaccagaacccgctgtttatactgaagacggtgcacgacctgaaaacatca	913
Dd	193 AGATCGGTGAGAAGGATGCCATTCTTGTGAATGGTGCCACCCTCTCCAACGTTA	252
Qy	914 ccattggcgacgacggcggcgatgtattcatcttttcaggctgatgccaaatatagacaatc	973
Dd	253 TCATTGTGTCTCCAGCGCGAGGGTGTTCCTCAAGGAACTTGCACCTCTCAACAATG	312
Qy	974 tgcaactcaccaactggttgagacgcgattaccgtttaagccaacacagcgcggcgcaaaa	1033
Dd	313 TCTGGTGGCGCATGTCTGTGAGGATCCTGTCACTCTCAAGCACAGCAGCGGAACCTCTT	372
Qy	1034 aatcccactgttaaatcactaacagttctcttcgacgacgctctgacaagatctcgcgc	1093
Dd	373 ACATCAACGGTGGTGGTCTTC-----CACGGCTCTGACAAGATCATTCAGT	420
Qy	1094 tgaatgccgatactaactgagcggttgacaacgtgaaggccaagcaagtcttggtactttg	1153
Dd	421 TCACGCGCGCGGAACCGTCCACGTCAAGGACTTCTACGCGGAGGACTACGGAAAGCTCT	480
Qy	1154 tacgactaacgcggcggtcacaacgggttaactggg	1186
Dd	481 CCCGACACTCGGNAACCTGCAAGGACACACGGTG	513
RESULT 14		
SC2G38		
LOCUS	SC2G38 40549 bp DNA BCT 13-DEC-2000	
DEFINITION	Streptomyces coelicolor cosmid 2G38.	
ACCESSION	AL445503	
VERSION	AL445503.1 GI:10803131	
KEYWORDS	ABC transporter ATP-binding protein; ABC transporter transmembrane protein; asnC-family transcriptional regulatory protein; hydrolase; integral membrane protein; lipoprotein; marG-family transcriptional regulator; oxidoreductase; oxidoreductase, iron-sulphur binding subunit; oxidoreductase, molybdopterin binding subunit; regulatory protein; secreted deacetylase; secreted lyase; secreted protein; tetR-family transcriptional regulator; two component system histidine kinase; two component system response regulator; ung, uracil DNA glycosylase.	
SOURCE	Streptomyces coelicolor.	
ORGANISM	Streptomyces coelicolor	
REFERENCE	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
AUTHORS	Redenbach,M., Kieser,H.M., Denapaitte,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.	
TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome	
JOURNAL	Mol. Microbiol. 21 (1), 77-96 (1996)	
MEDLINE	97000351	
REFERENCE	2 (bases 1 to 40549)	
AUTHORS	Saunders,D.C. and Harris,D.	
JOURNAL	Unpublished	
REFERENCE	3 (bases 1 to 40549)	
AUTHORS	Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-OCT-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park,	

Colney, Norwich, Norfolk NR4 7UH, UK

Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gcg, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2638. Location/Qualifiers
1. .40549
/organism="Streptomyces coelicolor"
/db_xref="taxon:1902"
1. .40549
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 2638"
complement(1..320)
/gene="2SCG38.01c"
complement(1..320)
/note="2SCG38.01c, possible regulatory protein(fragment), len: >106 aa; similar to N-terminal region of TR:CA01635 (EMBL:AL391072) Streptomyces coelicolor putative regulatory protein SC9A4.08, 116 aa; fasta scores: opt: 218 z-score: 308.8 E(): 9.3e-10; 46.2% identity in 91 aa overlap"
/codon_start=1
/transl_table=11
/product="putative regulatory protein(fragment)"
/protein_id="CAC13060.1"
/db_xref="GI:10803132"
/translation="MTAATAAGTRELPHPTRAEIRLEGVLHALSDPPVRLRIVRDLAADSGALFSCSHFDLPVTKSTTTTFRVLRSGVIRQTYRGTAKMNGLLRRDDLDIVFPGLLDALI"
misc_feature
1. .93
/note="nominal overlap with Streptomyces coelicolor cosmid SC2G4"
complement(326..329)
RBS
446..452
RBS
462..1541
gene
/gene="2SCG38.02"
462..1541
CDS
/note="2SCG38.02, probable oxidoreductase, len: 359 aa; similar to SW:NEMA_ECOLI (EMBL:D86931) Escherichia coli N-ethylmaleimide reductase (EC 1.-.-) Nema, 365 aa; fasta scores: opt: 509 z-score: 564.1 E(): 5.6e-24; 32.1% identity in 364 aa overlap. Contains Pfam match to entry

PF00724 oxidored_FMN, NADH:flavin oxidoreductase / NADH oxidase family"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase"
/protein_id="CAC13061.1"
/db_xref="GI:10803133"
/translation="MSALFEPFLRDTTIPNRMPPMCQVSAAPGSPAGVPGDMHF AHYGARVGGTGLIVVEATGVSPEGRIQPDGLWMDTQVEAFRITGLRSGQTVPA VOLAHAGKSTAQPMRGAPGVGDAYGWSALAPSAFAFRHPVPTLTVPDQGEAV GRDAARRALAAAGFEIAIEHGAHGYLHIFLSPHSNORTDAYGGSYANRTRELEV DAVRYWDDKPLFFRVSATDMLWEGGWTTPDTRVFARDLEAHGIDLDLDTSGGNVPR VRIPTGQTQVFFAARVAGSTLPVAAVGLITEPGCAEKILANGADAVLLGRELURN PSWAQHAARELGVDARMPDQYGWM"
468..1505
/gene="2SCG38.02"
/note="Pfam match to entry PF00724 oxidored_FMN, NADH:flavin oxidoreductase / NADH oxidase family, score 318.10, E-value 1e-91"
1739..2539
/gene="2SCG38.03"
1739..2539
/note="2SCG38.03"
/note="2SCG38.03, probable secreted lyase, len: 266 aa; similar to TR:Q04701 (EMBL:M94691) Fusarium solani pectate lyase A precursor (EC 4.2.2.2) Pella, 242 aa; fasta scores: opt: 605 z-score: 664.1 E(): 1.5e-29; 43.6% identity in 236 aa overlap. Contains possible N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted lyase"
/protein_id="CAC13062.1"
/db_xref="GI:10803134"
/translation="MTAVTRPRRAVTVGALGLSVGLMLTSCASSAQAATWPTPN GSEGVSTLSVSGTYDYGKMKRLYGTGDLGSGQDDEQPILEGAVLKNVILGAPA ADGVHCKGSCITLQNVWEDVGDAAATFRGSSSNVTVSGGAKGADKVPFGNAGT ITNIGDSSKKIVPCQYKICGNDGDGDPDSNGSGADGTYCKYSSSDITYK"
complement(2544..3350)
/gene="2SCG38.04c"
complement(2544..3350)
/note="2SCG38.04c"
/note="2SCG38.04c, conserved hypothetical protein, len: 268 aa; similar to TR:O53782 (EMBL:AL021943) Mycobacterium tuberculosis hypothetical 26.5 kDa protein MTV040.02, 240 aa; fasta scores: opt: 735 z-score: 848.5 E(): 0; 51.4% identity in 218 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC13063.1"
/db_xref="GI:10803135"
/translation="MGATVRAHVDERGAARMTCGDPGRLTLPRLSARSVLSLL QDEALPRLDRLVLVAGFDVGASTAARALSMAAGDLRDTDTGYRUSERLLERQR RLPDLGRVLAERLVSRGSPAVELAAARLPLVDWADTARALLVHVDARRRPAGLLTA FAAYVRLLADPVLPELLPDPWCTALRDYAVARYQREQSGQVRAHGPRT"
3399..5102
/gene="2SCG38.05"
3399..5102
/note="2SCG38.05, probable oxidoreductase, len: 567 aa; similar to SW:AIDB_ECOLI (EMBL:L20915) Escherichia coli oxidoreductase AIDB protein, 541 aa; fasta scores: opt: 1437 z-score: 1553.6 E(): 0; 45.8% identity in 517 aa overlap. Contains 2x Pfam matches to entry PF00441 Acyl-CoA_dh, Acyl-CoA dehydrogenase and match to Prosite entry PS00073 Acyl-CoA dehydrogenases signature 2"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase"
/protein_id="CAC13064.1"

THIS PAGE BLANK (USPTO)